

## DOCUMENT RESUME

ED 160 637

TM 007 724

AUTHOR Pilby, Nikola N.; Dishaw, Marilyn M.  
TITLE Construct Validation of Group-Administered Achievement Tests Through Individual Testing. BTES. Beginning Teacher Evaluation Study. Technical Note Series. Technical Note III-4. Phase III-A.  
INSTITUTION Far West Lab. for Educational Research and Development, San Francisco, Calif.  
SPONS AGENCY California State Commission for Teacher Preparation and Licensing, Sacramento.  
PUB DATE 25 Apr 77  
NOTE 129p.; For related documents, see ED 151 375 and 376 ; Parts of the appendices may reproduce poorly; Best copy available  
EDRS PRICE MF-\$0.83 HC-\$7.35 Plus Postage.  
DESCRIPTORS Academic Achievement; \*Achievement Tests; Elementary Education; Elementary School Mathematics; Grade 2; Grade 5; \*Group Tests; \*Individual Tests; Multiple Choice Tests; Reading Achievement; Test Construction; \*Test Items; \*Test Validity  
IDENTIFIERS \*Beginning Teacher Evaluation Study

## ABSTRACT

This technical note describes part of the work done to develop student achievement tests in reading and mathematics for use in a later phase of the Beginning Teacher Evaluation Study. While the achievement battery consists primarily of group-administered multiple-choice test items, several individual-testing studies were conducted to examine the validity of group-test measures and to gain insight for the improvement of group test items. The sample of students, data collection and analysis procedures, and results and conclusions are presented. The four individual-testing studies were: (1) Decoding skills in grade 2 reading; (2) Place value concepts in grade 2 mathematics; (3) Reading comprehension, grades 2 and 5; and (4) Testing formats in grade 5 mathematics. For each study, a rationale is given for the selection of particular tasks to investigate. Task procedures and materials are described and presented. Results are reported and conclusions are drawn about the appropriateness of group-test items and procedures. Items from each test are appended. (Author/CTM)

\*\*\*\*\*  
\* Reproductions supplied by EDRS are the best that can be made \*  
\* from the original document. \*  
\*\*\*\*\*

ED160637

SCOPE OF INTEREST NOTICE

The ERIC Facility has assigned  
this document for processing  
to:

TM

SP

In our judgement, this document  
is also of interest to the clearing-  
houses noted to the right. Index-  
ing should reflect their special  
points of view.

U S DEPARTMENT OF HEALTH,  
EDUCATION & WELFARE  
NATIONAL INSTITUTE OF  
EDUCATION

THIS DOCUMENT HAS BEEN REPRO-  
DUCED EXACTLY AS RECEIVED FROM  
THE PERSON OR ORGANIZATION ORIGIN-  
ATING IT. POINTS OF VIEW OR OPINIONS  
STATED DO NOT NECESSARILY REPRESENT  
OFFICIAL NATIONAL INSTITUTE OF  
EDUCATION POSITION OR POLICY

Technical Note III-4

CONSTRUCT VALIDATION  
OF GROUP-ADMINISTERED ACHIEVEMENT TESTS  
THROUGH INDIVIDUAL TESTING

by

Nikola N. Filby

and

Marilyn M. Dishaw

Submitted to

California Commission for Teacher Preparation and Licensing  
1020 "O" Street  
Sacramento, Ca. 95814

April 25, 1977

Far West Laboratory for Educational  
Research and Development  
1855 Folsom Street  
San Francisco, Ca. 94103

## PREFACE

This is one of a series of reports by the Far West Laboratory for Educational Research and Development, concerned with the study of teacher effectiveness. Under contract to the California Commission for Teacher Preparation and Licensing, with funding provided by the National Institute of Education, the Laboratory is studying teachers in second and fifth grade classes in order to identify teacher behaviors and classroom qualities that are related to reading and mathematics achievement. This multi-year research effort is titled the Beginning Teacher Evaluation Study (BTES). The study addresses a major policy issue of the California Commission which is the agency charged with certifying the appropriateness of teacher training programs throughout the State. Simply put, the Commission wants information about what teacher behaviors are related to student outcomes. This information will then be used jointly by the Commission and the State institutions that it certifies in order to better insure that beginning teachers receive training in areas that have been empirically demonstrated to affect student learning.

This report describes part of the work done to develop student achievement tests in reading and mathematics for use in the field study to be conducted during 1976-1977, known as Phase III-B of the BTES. These developmental activities were carried out in the year preceding the field study. The student achievement test battery developed for BTES consists primarily of group-administered multiple-choice items. This report describes several individual-testing studies conducted to examine the validity of group-test measures and to gain insight for the improvement of group-test items.

Several consultants assisted in the identification of important issues for study and in the development of individual testing methodology. Dr. James Wilson, Department of Mathematics Education, University of Georgia, consulted on mathematics measures. Dr. Roger Shuy, Center for Applied Linguistics, and Dr. Barbara Horvath, University of British Columbia, consulted on reading measures. The BTES staff gratefully acknowledges their helpful suggestions.

The BTES staff also wishes to thank Faye Mueller and Karen Ludwig for their help in data collection and scoring.

David C. Berliner  
Principal Investigator  
Beginning Teacher Evaluation Study  
April, 1977

## TABLE OF CONTENTS

|                                                                                                   |     |
|---------------------------------------------------------------------------------------------------|-----|
| PREFACE . . . . .                                                                                 | i   |
| INTRODUCTION . . . . .                                                                            | 1   |
| STUDY 1 - DECODING SKILLS IN GRADE 2 READING . . . . .                                            | 2   |
| Rationale . . . . .                                                                               | 2   |
| Description of Tasks . . . . .                                                                    | 3   |
| Design and Procedure . . . . .                                                                    | 7   |
| Results . . . . .                                                                                 | 8   |
| Summary of Conclusions . . . . .                                                                  | 13  |
| STUDY 2 - PLACE VALUE CONCEPTS IN GRADE 2 MATHEMATICS . . . . .                                   | 14  |
| Rationale . . . . .                                                                               | 14  |
| Description of Tasks . . . . .                                                                    | 15  |
| Design and Procedure . . . . .                                                                    | 18  |
| Results . . . . .                                                                                 | 21  |
| Summary of Conclusions . . . . .                                                                  | 29  |
| STUDY 3 - READING COMPREHENSION IN GRADES 2 AND 5 . . . . .                                       | 32  |
| Rationale . . . . .                                                                               | 32  |
| Description of Tasks . . . . .                                                                    | 33  |
| Design and Procedure . . . . .                                                                    | 34  |
| Results . . . . .                                                                                 | 35  |
| Summary of Conclusions . . . . .                                                                  | 43  |
| STUDY 4 - TESTING FORMATS IN GRADE 5 MATHEMATICS . . . . .                                        | 45  |
| Rationale . . . . .                                                                               | 45  |
| Description of Tasks . . . . .                                                                    | 45  |
| Design and Procedure . . . . .                                                                    | 49  |
| Results . . . . .                                                                                 | 50  |
| Summary of Conclusions . . . . .                                                                  | 60  |
| REFERENCES . . . . .                                                                              | 62  |
| APPENDIX A. Individual Testing Materials for Grade 2<br>Reading Decoding (Study 1) . . . . .      | A-1 |
| APPENDIX B. Group and Individual Testing Materials for<br>Grade 2 Mathematics (Study 2) . . . . . | B-1 |
| APPENDIX C. Group Testing Materials for Reading<br>Comprehension (Study 3) . . . . .              | C-1 |
| APPENDIX D. Group Testing Materials for Grade 5 Mathematics<br>(Study 4) . . . . .                | D-1 |

# LIST OF TABLES

| Table<br>Number |                                                             | Page |
|-----------------|-------------------------------------------------------------|------|
|                 | STUDY 1                                                     |      |
| 1               | Summary Data for Sight Words . . . . .                      | 8    |
| 2               | Summary Data for Word List . . . . .                        | 9    |
| 3               | Summary Data for Long and Short Vowel Nonsense Words . . .  | 10   |
| 4               | Context Clues Correlations . . . . .                        | 12   |
|                 | STUDY 2                                                     |      |
| 5               | Class Means by Subtest for Group and Individual Testing . . | 21   |
|                 | STUDY 3                                                     |      |
| 6               | Confusing Grade Two Items . . . . .                         | 38   |
| 7               | Confusing Grade Five Items . . . . .                        | 41   |
|                 | STUDY 4                                                     |      |
| 8               | Performance on Lines and Figures Items . . . . .            | 53   |

## INTRODUCTION

One part of the instrument development for Phase III-B of the Beginning Teacher Evaluation Study was the development and refinement of a set of student achievement measures in reading and mathematics. The conditions of a large field study necessitated the use of group-administered paper-and-pencil tests. A battery of tests was compiled, pilot tested, and revised, to increase reactivity to instruction (see Filby and Dishaw, 1975, 1976).

A supplementary part of the test refinement effort was a series of small studies using individual testing procedures to assess the validity of group-administered measures and to gain insights for improving test items. This Technical Note reports on four individual testing studies:

Study 1 - Decoding Skills in Grade 2 Reading

Study 2 - Place Value Concepts in Grade 2  
Mathematics

Study 3 - Reading Comprehension, Grades 2 and 5

Study 4 - Testing Formats in Grade 5 Mathematics

For each study, a rationale is given for the selection of particular tasks to study. Task procedures and materials are described. Results are reported. Conclusions are drawn about the appropriateness of group-test items and procedures. Actual revisions in the test battery are recorded in the full Technical Report on test development (Filby and Dishaw, 1976).

## STUDY 1

### DECODING SKILLS IN GRADE 2 READING

#### Rationale

The most direct measure of a student's ability to read is to ask him to read aloud. As the student reads, the characteristics of his responses reveal the nature of his word attack strategies. Unfortunately, in group testing this direct measure is impossible to obtain. In group testing the student must make a written rather than an oral response. Such a written response task is a more indirect measure of decoding skill than the task of actual word reading. The purpose of this component of individual testing is to validate the written decoding tasks in the group test battery.

The basic validity question is whether students are rank ordered consistently from one method of assessment to another. That is, students with the best performance in reading words aloud should receive the highest scores on written group test measures; students who perform relatively poorly on word reading should perform relatively poorly on other measures as well. To the extent that this is true, the more indirect written measures are acceptable substitutes for oral assessment.

A secondary concern is the absolute level of performance on different measures. Two competing factors could influence task difficulty: (1) the written measures used in the group test battery are multiple-choice items. The student must recognize and select the correct word. In contrast to this recognition process, actual word reading is a production process. Recognition is generally an easier process than production. For this reason, scores might be higher on the group test; (2) individual testing

provides a testing situation which elicits the best from the child. The interviewer's attention is directed solely to the one student, who knows his performance is being watched continuously and who is encouraged to do his best. If students pay attention more and try harder in individual testing, scores might be higher in individual testing.

The most interesting case of discrepancy in task difficulty is that in which scores are lower on the production task in individual testing. Ability to decode common patterns advances rapidly in grade 2, and performance reaches ceiling in group tests of specific decoding skills. Such ceiling effects create a problem for data analysis. Shifting to a production task might extend the upper range of the scale. Since teachers presumably use production (word reading) as a criterion, we might need the production measure to capture the full effect of instruction.

#### Description of Tasks

Four separate tasks were included in this study. Each involved oral reading of words or sentences, and was parallel to a component of the group test battery. Materials for individual testing are shown in Appendix A. Tasks are described in the order in which they were administered to students.

Task 1 - Sight Words. One subtest in the group test battery involves the recognition of frequently used function words. Since these are words students most often memorize or learn to recognize automatically, the label "Sight Words" is used for this subtest. The testing procedure is that the tester pronounces a word aloud and the student must find that word in a set of five choices. For example, the word "where" spoken by

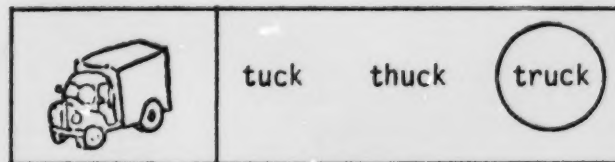
the tester, must be located and circled in the set of words below:

where      whate      might      were      whose

This group administered subtest has 14 items, with one target word repeated twice in different sets.

The parallel, individually administered production task is to ask the student to read aloud each of the 13 target sight words (see Appendix A).

Task 2 - Word List. A more frequently used format in the decoding subtests of the group test battery is to present a picture and ask the reader to select the one of three words which correctly names the picture. The response choices are identical except for a specific feature. A sample group-test item is the following:



The focus of attention in this item is on decoding initial consonant blends. Other items focus on single consonants, long vowels in a final "e" or digraph pattern, vowel diphthongs, vowels followed by "r," or complex multi-syllabic patterns. In all, there are seven subtests with a total of 90 items which use this test format.

The corresponding individual testing task is for the student to read aloud the words which are correct responses in the group test items (such as "truck"). A subset of 22 of the 90 possible words was selected for oral reading. This subset includes at least two instances of each of the decoding elements that are the focus of the seven subsets.

Task 3 - Word List in Context (Context Clues). In the previous task words are read in isolation. The student must rely totally on decoding (or word memory) skills. Most natural reading involves reading words in context (i.e., sentences or paragraphs). When reading connected prose, the reader can use information from the surrounding context to help decode any one word. This use of information from the surrounding context is sometimes termed "context clues." It is presumably a reading skill distinct from knowledge of phonetic decoding principles.

Individual testing provides an ideal opportunity to assess the reader's ability to make use of context clues. While performing Task 2, the reader will presumably miss some words. The student can then be asked to read those same words in sentence context. The tester can provide help to the student in reading non-target words, so that in all cases the student will have a context in which to fit the target word. The number of additional words the student can read within a sentence context provides a measure of ability to use context clues. The sentence contexts for each of the 22 Word List items are shown as Task 3 in Appendix A.

No directly parallel measure is possible in group testing, for two reasons: (1) The tester cannot provide help with words that are part of the context, so the measure becomes one of reading context rather than using context; and, (2) even if the student succeeds in reading and using context, he must still select the correct response from among the alternatives given. By careful comparison of two group-test scores, it is possible to overcome objection 2; but there is no ready solution to objection 1.

Compare the two items below:

|                                          |                                                                                   |       |      |       |
|------------------------------------------|-----------------------------------------------------------------------------------|-------|------|-------|
| Item A -<br>Decoding<br>Words<br>(group) |  | swing | sing | sting |
|------------------------------------------|-----------------------------------------------------------------------------------|-------|------|-------|

|                                         |                                        |      |       |  |
|-----------------------------------------|----------------------------------------|------|-------|--|
| Item B -<br>Context<br>Clues<br>(group) | The _____ on the playground is broken. |      |       |  |
|                                         | swing                                  | sing | sting |  |

Item A is a measure of decoding consonant blends. Item B is an example of a context clues item from the group test. To answer item B, the student must read the sentence, infer the missing word, then decode blends to select the answer. Subtest B will be more difficult than subtest A. A difference score comparing performance on these two measures holds constant the decoding factor and estimates the student's ability to read and use sentence context. Theoretically, this difference score would be the group measure most similar to the individual testing improvement score. It still involves more word reading than the individual measure.

Task 4 - Long and Short Vowel Nonsense Words. Two of the group-administered subtests described in Task 2 focus on decoding common long vowel patterns - long vowels with silent final "e" and vowel digraphs. Sample group test items of these two types are shown below:

|                                                                                     |      |     |      |
|-------------------------------------------------------------------------------------|------|-----|------|
|  | kute | kit | kite |
|  | poal | pal | pail |

These long vowel subtests are of particular interest because long vowel patterns are frequently taught in grade 2. For this reason, a separate list of long and short vowel words was included in individual testing. In this word list, rather than using the real words that are correct responses in group test items, nonsense words were used. Using nonsense words provides a more pure measure of the ability to generalize phonetic principles. The child might simply have memorized the word "kite," although to distinguish "kite" from "kit" the child must have internalized the critical feature of the final "e." The child certainly has to go beyond rote memory to read a word like "kipe," which he has never seen before. In so far as one would like to interpret the group tests as measures of the child's understanding of a specific phonetic generalization, the nonsense words test is perhaps the most valid criterion. It is also probable that scores will be lower when nonsense words are used; and this might help eliminate ceiling effect problems, as discussed previously.

The two group-administered subtests have a total of 22 items. In eight items the correct response is a short vowel word. In the remaining items, the correct word has one of the following long vowel patterns: "a-e," "i-e," "o-e," "ai," "ea," "ee," or "oa." The nonsense words list is an extension of a list used by Robert C. Calfee at Stanford University. It covers all long vowel patterns, with matched short vowel words (see Appendix A).

#### Design and Procedure

All tests were administered in March of 1976. The group tests were administered as part of the regular testing sequence in Phase III-A

Extension of the Beginning Teacher Evaluation Study. Individual tasks were administered two weeks prior to the group testing.

Twenty students participated in individual testing, ten from each of two grade 2 classes. In one class, a few students were unable to read more than a few words and experienced extreme frustration with the task. These students were replaced. With this exception, the full range of grade 2 reading levels was included in the sample.

### Results

Sight Words. The sight words task compares oral reading of basic function words in individual testing to group-administered 5-choice items with an oral stimulus. Performance in these two conditions is summarized in table 1.

TABLE 1  
Summary Data for Sight Words

| Test Format | # of Items | Mean Correct | High Score | Low Score | Correlation |
|-------------|------------|--------------|------------|-----------|-------------|
| Group       | 14         | 72%          | 100%       | 13%       | .81         |
| Individual  | 13         | 84%          | 100%       | 54%       |             |

Scores on this subtest were significantly higher in individual testing than in the group multiple-choice format. This increase was caused by a dramatic upward shift at the bottom end of the distribution. Low score on the group test was 13% correct; low score in individual testing was 54% correct. This shift is greater in magnitude than for other tasks discussed later. This result indicates that the group test underestimates the performance of students at the low end of the scale.

One possible interpretation is that the oral stimulus is harder for low ability students to process than the pictorial stimulus. The student hearing a word aloud must translate the word into his own internal language, hold the word in memory, and locate the graphic form. The pictorial representation allows the student to form his own verbalization, with the picture as a constant reminder.

Another interpretation is that the number of choices is a critical factor. With five words to choose among, the low students may compare words less carefully and make quick, erroneous responses.

Regardless of interpretation, this test seems valid as indicated by the correlation, but it underestimates the absolute level of performance for low ability students.

Word List. This task compares oral reading of words to group-administered 3-choice items with a pictorial stimulus. Performance in these two conditions is summarized in Table 2.

TABLE 2  
Summary Data for Word List

| Test Format | # of Items | High Score | Low Score | Correlation |
|-------------|------------|------------|-----------|-------------|
| Group       | 90         | 89%        | 18%       | .92         |
| Individual  | 22         | 100%       | 14%       |             |

This task produced a good range of scores in both conditions. Partly as a result of this range, there was a high correlation between scores on the group-administered and the individual measures. The group test appears to be a valid measure of decoding skill.

Interpretation of differences in mean correct cannot be made clearly. The 22-item subset of words used in oral reading represents the range of decoding patterns included in the full set of 90, but it was not matched to the total in terms of overall difficulty level. Nevertheless, it is clear that the underestimation of low scores on the group test of sight words is not repeated in this group test using a different format.

Long and Short Vowel Nonsense Words. This task compares oral reading of nonsense words with long and short vowel patterns to group-administered, 3-choice items with a pictorial stimulus, where the choices vary in vowel pattern. In individual testing only decoding of the vowel was considered in the scoring of the items. Errors in surrounding consonants were ignored. Performance in the two conditions is summarized in Table 3.

TABLE 3  
Summary Data for  
Long and Short Vowel Nonsense Words

| Test Format | # of Items | Mean Correct | High Score | Low Score | Correlation |
|-------------|------------|--------------|------------|-----------|-------------|
| Group       | 22         | 68%          | 100%       | 11%       | .70         |
| Individual  | 26         | 56%          | 88%        | 27%       |             |

In this task, unlike the two previous ones, performance was poorer in individual testing than on the corresponding group-administered items. Generalization of a decoding principle to nonsense words is somewhat independent of and generally harder than selecting real words. In the group test four students scored 100% correct. On the nonsense words, the high score was 88%. Several students who achieved high scores on

the group test did quite poorly on the nonsense words.

These results suggest that an oral reading task with nonsense words could be a useful extension of the BTES test battery. This measure is relatively more independent of existing measures than other individual tasks. It also lowers the range of scores and thus reduces ceiling effect problems.

One could argue that the nonsense words extend the range of patterns tested and that these additional, harder items lower top scores. This cannot be the whole explanation. Even on a subset of items which match the group test ( $\bar{a}$ ,  $\bar{i}$ ,  $\bar{o}$ ) some students scored quite well on the group test and poorly on the nonsense words.

Context Clues. In individual testing, use of context was assessed through a comparison between performance on Task 2, word reading, and performance on Task 3, words in context. Of the 20 students tested individually, 5 made only 2 or fewer errors on word reading and therefore had no opportunity to show use of context. These 5 students were deleted from all context clues analyses. For the remaining 15 students, ability to use context was expressed as a percent improvement score. If a student missed 10 words in Task 2 and was able to read 6 of the 10 words with the help of sentence context, he received a score of 6 over 10, or 60% improvement. The mean improvement score was 62%, with a low of 38% and a high of 100%.

On the group test, two scores were calculated. One was the number of items correct (corrected for guessing) on fill-in-the-blank items like item B, shown previously. This is referred to as the "raw" context clues score. The second score was a difference score removing the effect of

ability to decode blends. The formula used was:

$$\text{Context Clues Adjusted} = \text{Raw Context Clues} + (9 - \text{Decoding Blends})$$

It was expected that compared to the raw score the adjusted score would be less highly correlated with overall decoding scores and more highly correlated with the percent improvement score. The question of primary interest was whether the group adjusted score would correlate highly enough with the individual measure to be an acceptable substitute.

TABLE 4  
Context Clues Correlations

| Variable                      | Test Format | Correlations |     |     |     |
|-------------------------------|-------------|--------------|-----|-----|-----|
|                               |             | 1            | 2   | 3   | 4   |
| 1 Decoding Words              | Group       |              |     |     |     |
| 2 Context Clues-Raw           | Group       | .76          |     |     |     |
| 3 Context Clues-Adjusted      | Group       | .37          | .82 |     |     |
| 4 Decoding Words              | Individual  | .93          | .67 | .24 |     |
| 5 Context Clues-% Improvement | Individual  | .46          | .64 | .59 | .46 |

Correlations are shown in Table 4. These figures indicate the following:

1. The adjusted group context clues score is much less highly correlated with decoding than is the raw score (as predicted).
2. The adjusted group score is also slightly less highly correlated with the individual percent improvement score (contrary to prediction; probably no real difference between the two correlations).

3. Neither group measure of context clues correlates especially highly with the individual measure.

It appears that the percent improvement score assesses a skill moderately independent of decoding but which cannot be fully captured in group measures.

#### Summary of Conclusions

Encouraging results were obtained on the validity of the group-administered decoding subtests. The Sight Words subtest, using an oral stimulus, and the specific feature subtests, using a pictorial stimulus, all correlate highly with performance on reading words aloud. These high correlations justify the use of group measures of decoding skill. It should be noted that the Sight Words subtest underestimates performance, particularly for low ability students, so the absolute value of scores should be interpreted with caution.

The use of nonsense words in individual testing lowered performance levels and slightly reduced the correlation with the group measure. If individual testing could be done in Phase III-B, this would be a profitable area. A nonsense words task would reduce ceiling effect problems and would provide a measure of transfer.

The context clues measure in the group test did not hold up very well. A score adjustment can be used to reduce the role of specific decoding skill. Neither the raw nor the adjusted score, however, correlated very highly with the individual testing measure. Unless other indicators such as reactivity are more promising, the context clues measure should be dropped.

## STUDY 2

### PLACE VALUE CONCEPTS IN GRADE 2 MATHEMATICS

#### Rationale

During Phase III-A of our study we encountered classrooms where a wide variety of concrete materials were available to support an individualized or small group instructional setting for students in mathematics. Specifically, we noted that manipulative materials were being used as a major tool in teaching place value concepts. Although our group test included place value items with pictorial representation these items are still an abstraction of the tasks a student may have experienced within a manipulative based program. Individual testing was therefore viewed as an excellent opportunity to investigate place value concepts with concrete materials and to obtain information on the appropriateness of the group test format. It was also theorized that, for students who operate primarily within an individualized setting rather than a group seatwork setting, the group test may be underestimating their skill level due to their unfamiliarity with a solely pencil-and-paper seatwork situation. In conjunction with the focus on place value, items on numerals were included since a student's ability to order and compare numerals involves an understanding of place value.

While we were at it, selected items from the group test scales on Missing Addends and Equalities and Inequalities were included to provide additional information for the already anticipated modification of these items. Finally, a few addition problems involving regrouping were included out of a clinical interest in the transference and application of place value concepts.

### Description of Tasks

The group test, administered prior to individual testing, contained six subtests from the 1975-76 BTES pilot battery. Appendix B includes a copy of the group test and item numbers are circled to indicate those items that were tested individually. A copy of the individual tests is also included in Appendix B.

The five tasks included in the individual testing are described in the same order in which they were presented to the students.

Task 1 - Numerals. Using the same page and initial procedure as in the group test, each question was read to the student. The student was asked to circle his answer and explain his choice. For items 1, 2, and 4 the children were then asked to order (read) the answer choices from the smallest to the largest. On item 3, they were asked to state if each number came before 55 or after 62. Particular attention was given to items 2 and 4 which, in contrast to the other three items, showed little or no change in p values (percent of students passing an item) between the "A" and "B" test administrations.

Task 2 - Equalities and Inequalities. Using the sample items from the group test, the students were asked what each of the signs meant and to work the sample item. If the signs were defined or used incorrectly then the interviewer corrected the student's definition prior to his solving the test items. The test items were presented in an open-ended format and the following procedure was used.

- a. The student read the number sentence.
- b. Was asked what he would do first in solving this problem.
- c. If the student indicated that he would solve each side of the sentence, he was asked to do so above each part and then to select the sign and explain his choice. Children were also allowed to solve it first and explain later. Completed solutions were read by the students.

Our interest in investigating these items stemmed from the fact that there was a high no response rate at both the "A" and "B" test points. The students may not have been responding because they didn't know the terminology (symbols), were frustrated by the number of steps required to solve each problem and therefore attempted only a few items, or felt the layout to be overwhelming and thus "gave up" at the beginning.

Task 3 - Missing Addends. Three items from the group test were presented in an open-ended format. For the first two problems the following procedure was used.

- a. The student read the equation.
- b. The student solved the problem writing his/her answer in the designated blank.
- c. The student read the completed equation.

The last item was presented in two ways:

1.  $44 = \square - 3$

(group test  
format)

2.

$$\begin{array}{r} \square \\ - 3 \\ \hline 44 \end{array}$$

The following procedure was used.

- a. The student read both problems.
- b. The student was then asked if both the problems were asking the same question.
- c. The student solved whichever problem he/she thought was easier.
- d. The student read the completed equation.

If a student had difficulty in understanding a problem, manipulatives were used to illustrate it and/or the problem was rewritten in an effort to help the student reach a correct solution. The first and third items in this set are ones which were presented in the  $a = \square - b$  format and in the group tests they did not discriminate between the more able and less able students. Individual testing should provide information on why some of the more able students are having difficulty.

Task 4 - Place Value. Children were given the choice of using the manipulatives provided by the interviewer or those available from their classroom. The tasks are parallel to items 7, 12, 13, and 18 from the group test. Specific tasks were as follows:

1.
  - a. The interviewer asked the student to write the number 63.
  - b. The interviewer then asked the student which number was in the one's place and to draw a circle around that number.
  - c. The student was then asked to show 63 with manipulatives.
2.
  - a. Using manipulatives (sticks), the interviewer placed 3 tens and 4 ones on a table while stating the same.
  - b. The student was asked what number 3 tens and 4 ones is.
  - c. The student was instructed to write his/her answer to step b.
3.
  - a. Without stating how many, the interviewer placed 2 hundreds, 4 tens, and 3 ones on a table.
  - b. The interviewer asked what number matched the whole group of sticks.
  - c. The student then wrote the answer to step b.
  - d. The interviewer checked/validated the student's answer by having the child match his/her number to the sticks. That is, starting with the written numeral, the interviewer asked which number is in the one's place. Then asked how many ones were on the table. The interviewer repeated this procedure for tens and hundreds.
4.
  - a. The interviewer placed 3 hundreds on a table and then asked how many hundreds there were. The student was then asked to write the numeral for three hundreds.
  - b. The interviewer explained that some tens were going to be added to the hundreds. Placing a card with a plus sign (+) next to the pile of hundreds, the interviewer instructed the student to write "+" after his/her "300."
  - c. The interviewer then placed 6 tens down after the plus sign, and asked how many tens there were. The student was asked to respond and then to write the numeral for 6 tens.
  - d. The interviewer placed another plus sign and then 8 ones and repeated the procedure in step c.

4. e. The student was then asked to read his statement. The interviewer asked the student if he/she could write these numbers a shorter way without any plus signs. To write it as all one number that would mean the same thing.
- f. If the solution was correct, the student was asked why the zeros dropped out. "What happened to the zeros? Why don't you need to write them in the answer?"

Task 5 - Addition with Regrouping. Three items from the group test were used and depending on the student's ability he/she was guided through as many of the following steps as possible.

- a. The student read the problem.
- b. The student solved the problem in written form.
- c. The student explained the solution, and was asked why he/she had to carry.
- d. The interviewer then probed the child on whether he/she was carrying a one or a ten. In some cases the student was asked to write the sum of the ones column and to explain which digit was carried and if that was a one or a ten. He/she was also asked if that digit remained a one or a ten after it was carried.
- e. The interviewer then had the student demonstrate his/her solution, to one or more of the problems, with manipulatives.
- f. After proving a solution with the manipulatives, the student was asked if he/she could get the sticks to look the same as the written number. That is, "How can you change this (5 tens and 16 ones) to be the same as your written answer (6 tens and 6 ones)?" As effort was made to see if the student exchanged ten ones for one ten.

#### Design and Procedure

Prior to the individual testing, two classes were administered a group test which included six subtests or partial subtests totaling sixty items. Each subtest was presented in the same way as in the 1975-76 test battery. Conditions were similar to those for a regular group test session: one booklet, 45 minutes, 3-5 minute break between the two sections. From the group test results children were selected for

individual testing. Particular attention was given to selecting students who had done fairly well on the total test but had missed a number of the items which had been selected for individual testing. The group tests were administered during the last week of February, 1976 and the individual testing was conducted during the first two weeks of March, 1976. Both the group and individual tests were administered by the same person. Prior to testing the students individually, a practice session was held with students from a non-study classroom. After that session, some procedural modifications were made and a coding form was developed. Appendix A includes a copy of this coding form.

Sample. The students were selected from two classrooms both of which had participated in the Instructional Time Study (October to December 1975). They were selected because their math programs were widely different and it was known through observation that in one class the concept of place value had been taught primarily with manipulative materials. For this report they have been labeled as class A and Class B.

Class A provides a traditional instructional setting with seatwork in large groups and paper-and-pencil oriented materials from workbooks, the Modern School Mathematics text, and teacher prepared dittos. Class B is individually structured and provides small group instruction. The math centers employ the following materials: centimeter rods, number blocks with task cards, bean and beansticks, unifex cubes, E.S. pattern blocks, Veri-Tech math program, chip trading games, and a number of teacher made math games. Eighteen students from class A took the group test and ten were selected for individual testing. Twelve students from class B took the group test and seven were tested individually.

Materials. Popsicle sticks were bundled into groups of hundreds and tens and the single sticks were used as ones. Since the children from class A were not familiar with manipulatives a short explanation of their use was given. Pictorial items 5 and 6 from the group test were used as reference points and illustrated with the popsicle sticks. For example, item 5 shows four bundles of ten sticks each and three single sticks and asks "How many TENS are there in this picture?" The student was asked to answer that question and then shown how the test materials matched the picture. Ones and tens were discussed and then, with student participation, item 6 was illustrated. This discussion of the materials was conducted immediately prior to the place value task.

Procedure and Scoring. The individual sessions were all audiotaped and each question was either read to the student or by the student. Not all the steps outlined previously were considered in grading an answer as right or wrong. For data purposes, only those steps which were required by the group test item were graded. Non-graded steps included such things as illustrating an addition problem with manipulatives and explaining why the zeros dropped out from expanded to compact notation. Although the responses to the non-graded steps are not included in the score data they are discussed in the results section. Answers to the graded steps were considered right or wrong on the basis of a student's solution prior to prompting or discussion. In most cases, the interviewer worked with the child in an effort to help him or her reach a correct solution. In this way a better insight could be gained on a child's level of skill and his ability to apply that skill or concept. If, on a particular item, most of the children failed to understand it even with

guidance this would indicate that they lacked the skills necessary to solve the problem and its appropriateness would be in question.

### Results

Individual testing revealed some major differences between the students from class A and those from class B. Table 5 summarizes the results separately by class and subtest (task).

TABLE 5  
Class Means by Subtest for Group  
and Individual Testing

| Subtest                  | # of Items | Class A (N=10)     |                         | Class B (N=7)      |                         |
|--------------------------|------------|--------------------|-------------------------|--------------------|-------------------------|
|                          |            | Group Mean Correct | Individual Mean Correct | Group Mean Correct | Individual Mean Correct |
| Numerals                 | 5          | .48                | .50                     | .49                | .60                     |
| Inequalities             | 3          | .64                | .90                     | .48                | .68                     |
| Missing Addends          | 3          | .53                | .20                     | .24                | .19                     |
| Place Value              | 4          | .75                | .75                     | .43                | .82                     |
| Addition with Regrouping | 3          | .60                | .77                     | .29                | .76                     |
| Total Test               | 18         | .60                | .62                     | .39                | .61                     |

Results are reported for each task both in terms of the responses across students and between classes. Item percentage values given within the following task discussions are based on all 17 students unless stated otherwise.

Numerals (5 items). Individually the children were first asked to answer the question and then asked to order the numerals. On items 1, 2, and 4, 90 percent of the students who solved the question correctly were able to order the numerals correctly. The ordering on item 3 ("Is it before 55 or after 62?") proved to be more difficult with only 50 percent of the students who had answered the item correctly being able to do this task. On items 1, 3, and 4, where the process of ordering was directly related to the process required by the problem, 71 percent of the students who were initially incorrect, but were able to order correctly, recognized and identified their errors while in the process of ordering.

Being able to order correctly did not help any of the children recognize their errors on item 2. The most frequent incorrect responses to this item were 900 and 899. The children who gave these responses were more likely to explain their answer by saying that it "sounded right." The students who were correct were more likely to explain their solutions in terms of a specific process. Many focused on the numeral 89 only and others used the starting point of 887 or 888 to count up to 890.

The explanations given by the students on item 5 also reflected conceptual differences that were appropriate per the item's purpose. The more able students indicated their realization that 1,000 is greater than 100 and therefore all they had to do was figure out which choice had the most thousands. Although most of these children did not know how many hundreds were in a thousand they appeared to work under the assumption that there wouldn't be more hundreds in the picture than were needed to make one thousand. The explanations given by the students who were incorrect showed that they did not fully understand that one thousand was

greater than one hundred nor why that might be important. Some counted the 100's and 1,000's as equal and ended up differentiating between groups of ten; others focused only the amount of hundred's within each figure; and some ignored the values and simply counted containers.

Performance on the individual task was slightly better than on the group test with class B showing the greater gain. The items did test the child's knowledge of the number system and his/her ability to order two and three digit numerals. Although items 2 and 4 were difficult they appeared to be appropriate.

Equalities and Inequalities (3 items). Individual testing produced comparable gains in performance for both classes. For class A and B respectively, there was a 41 percent and a 42 percent increase in the mean proportion correct. These gains can partly be attributed to the terminology being defined better for the students. Although the same sample item was available as a "key" in the group test, more attention was given to it in the individual session and to each child's definition of the signs. The children were as likely to reverse the meaning of the signs ( $>$ ,  $<$ ) as they were to identify them correctly (41% each). The children were also required to read each sentence outloud before solving it and this task probably contributed to the increased p values.

Once the students understood the terminology, they were able to apply a step-by-step process in solving the problems. Indeed nine out of the ten students from class A easily applied the process of evaluating both sides of the sentence, determining the relation between the two, and then selecting the sign which reflected that relationship. A few students referred back to the sample item before selecting the symbol.

Eighty-four percent of the students were correct on item 6 and most of them had to solve each side computationally before recognizing which was greater. On item 7 most of the children were immediately able to recognize that it contained two sets of ten without having to compute. This item had a p value of .90. Item 8 was the most difficult but still 65 percent of the students were able to solve it correctly. With a few exceptions, these students were able to recognize the need for the plus sign without knowing the sum of  $8 + 5$ . Of the six children who were not able to solve this problem correctly: two were confused by the change of signs and tried to continue with  $>$ ,  $<$ ; two tried a minus sign; one said she didn't know how to do that type of problem; and for one student the item was skipped.

It has been proposed that in future tests a symbol key be included for this scale. Considering the children's confusion over which sign is which that might be useful but the same purpose could be met by using an additional sample item and instructing the children to refer to it if necessary. The layout of the test page is probably overwhelming for some and the items should be separated onto two pages according to the signs being used. Redesigning the scale layout, including additional and better sample items, and providing better instructions should decrease the non-response rate and increase the appropriateness of this scale.

Missing Addends (3 items). On the group pretest class B's performance was at the chance level and dropped slightly with individual testing. Class A had done better on the pretest with a mean correct of 53 percent. However, with individual testing class A's performance decreased considerably and there was little difference between the two classes.

These items were quite difficult for the children as was aptly illustrated by their reaction to item 9 ( $1 = \square - 6$ ). Twenty-four percent read this item in reverse ( $6 - \square = 1$ ) and even when they read it correctly they overwhelming (by 82%) solved it in reverse. In responding to their solution of  $1 = 5 - 6$  the interviewer confirmed that this solution was correct if the problem was read from right-to-left. When asked if the same answer was correct if they read the problem from left-to-right, most recognized that it was not. A number of minutes were spent trying to help the students figure out the correct solution. The question was orally rephrased, manipulatives were used and the problem was rewritten in vertical form. However, few of the children were able to revise the framework they had started with and visualize  $\square - 6 = 1$  as a new and separate problem. It appeared that once they thought of it as  $6 - 5 = 1$  they couldn't change their viewpoint. Only one student was able to solve this item correctly without prompting or discussion.

The length of time spent discussing the first item probably helped some students and confused others in their approach to the last two items. Thirty-five percent were correct on item 10 and each of these students indicated that they had added the two known numbers to find the unknown. Students in class A were able to explain this process by stating that the teacher told them that was how to solve the problem, i.e., "teacher says that when you have a blank minus something equals something you add the two somethings together." Students in class B were not able to give a reason for applying the inverse relationship.

The last item fared somewhat better than the first with 18 percent of the students solving it correctly. As stated in the task section this item was presented horizontally and vertically. Nearly 62 percent picked

the vertical format as easier to solve. Most of the students recognized that the two problems were essentially the same. Although only 18 percent solved it correctly a total of 41 percent approached it correctly but used the marking method to solve it and obtained 46 or 48 for an answer.

The drop in p values for class A may have been due, at least in part, to the open-ended format in the individual test. Without the clues provided by the multiple-choice answers the students may have been more likely to respond with the obvious solution. Also, success on these items for class A students seemed to depend on whether or not they remembered the "rule" previously taught by their teacher. Perhaps on the group test more of them remembered that rule.

In any case, the fact that the children had a difficult time understanding items 9 and 11 even with guidance indicates that such items using the format of  $a = \square - b$  are conceptually more difficult than those where the missing value is in an alternate position. In the future the known sum or difference will be placed in a terminal position to the right of the equality sign and items will be presented in both a vertical and horizontal format.

Place Value (4 items). Here we see a dramatic difference between class B's individual and group test scores. While class A's performance remained the same at 75 percent correct, class B experienced a 91 percent increase by going from a mean p of .43 to .82 on the individual test. The children from class B had no problem adjusting to the test materials and only one student chose to use materials from the classroom. Five students were scored as correct on all the items. Of these five students three were also able to explain why the zeros dropped out from compact

to expanded notation (e.g., while pointing to 300: "Cause you add the 8, and then the 6 in the tens."). The two students who had some difficulty were not able to go from expanded to compact notation correctly and were either unable to recognize the number represented or were unable to write it correctly.

The students in class A needed some discussion and illustration before adjusting to the materials and some had difficulty illustrating the first problem. Only five out of the eight students, who had completed steps a and b correctly, were able to illustrate it correctly. Recognition was not a problem with the last three items although the children were more likely to count the total group for their answer than the students from class B. Forty percent solved all four items correctly on both the group and individual test. Item 15 was the most difficult with four students being unable to go from expanded to compact notation correctly. Of the six who were able to complete step 3 (compact notation) correctly, two had written the expanded form incorrectly (e.g.,  $300 + 600 + 800 = 368$ ). The students who were incorrect on items 13 and 14 tended to count the tens as ones or the ones as tens and arrived at such answers as 70 and 207.

Considering the responses from class A and B separately we might conclude that for children who learn place value concepts primarily with manipulative materials in an individualized small group setting, individualized testing with manipulatives is more fully reflective of their knowledge. Scores obtained from pencil-and-paper group testing may be negatively influenced by how far a particular mathematics program departs from a pencil-and-paper seatwork situation. This would explain the discrepancy between class B's group and individual test scores.

Addition with Regrouping (3 items). Again there is a large gain for class B from the group to the individual testing. There was a 162

percent increase ( $\bar{p}$  .29 to .76) from the group to the individual test. The children in class B were more able to explain and illustrate the problems than the students in class A. None of the seven students in class A who got all three items correct seemed to know that they were carrying a ten versus a one whereas two out of the five students in class B who solved all the problems correctly showed a clear understanding that they were carrying a ten and not a one. Children from class B were also more able to exchange ten ones for one ten when illustrating their solutions to the regrouping problems. The five students from class B who were correct on all the regrouping problems were the same five students that were correct on all the place value problems. Looking at all seven items on place value and addition with regrouping together, these five students went from a mean of 40 percent correct on the group test to 100 percent correct on the individual test.

The nuances, hesitations, questions, and self-corrections made by the students while in the process of solving the place value and addition problems were very informative. Although the children from class A did quite well, their performance appeared to be more mechanical than that of the children from class B. The children from class B appeared to have a better conceptual understanding of place value and its application than the children from A.

Why then did the students from class B do so poorly in addition with regrouping on the group test? Testing conditions did not appear to be the source of variance. Although there was a short interruption towards the end of the group test and the children's attention was somewhat less toward the end than at the beginning, these factors are not uncommon in testing and were not considered sufficient to explain the discrepancy that occurred.

Another possible explanation is that the students were unfamiliar with short form regrouping. Possibly they were more familiar with an instructional algorithm such as one shown in the following example.

$$\begin{array}{rcllcll} 49 & = & 4 \text{ tens} & + & 9 \text{ ones} & = & 40 & + & 9 \\ +32 & = & \underline{3 \text{ tens}} & + & \underline{2 \text{ ones}} & = & \underline{+30} & + & \underline{2} \\ & & 7 \text{ tens} & & 11 \text{ ones} & = & 70 & + & 11 & = & 81 \end{array}$$

If their instruction had emphasized instructional algorithms they may not have been able to make the transfer to the short form on the group test. However, it's quite possible that in individual testing the emphasis on place value and the sequence in which it was presented may have influenced their performance on the regrouping problems. These students were familiar and comfortable with the place value tasks and received positive feedback on their performance. Their success and the emphasis given to place value may therefore have led these students to recognize the need for regrouping in the addition problems which immediately followed the section on place value. If that was the case, 76 percent may very well be an overestimate of their short form regrouping skills and the 29 percent on the group test may be an underestimation.

#### Summary of Conclusions

Reviewing the summary data presented in Table 5, class A's total test scores were stable from the group to the individual testing. Class B's performance increased for the total test but most of the increase resulted from the scores on the last two subtests, place value and addition with regrouping. This trend of performance on the first three subtests being more consistent across test situations than for the subtests on place value and addition with regrouping is further supported by the correlational data. A group-test total score in Numerals, Equalities and

Inequalities, and Missing Addends correlated .70 with the corresponding 3-subtest total from individual testing. This correlation still reflects a fair amount of variability. However, a high correlation would not be expected since the selection of these items was based on previous data which indicated that these items were confusing for some students. Item difficulty plus format changes in the individual testing account for some of the variability. A composite score on place value and addition with regrouping yielded a correlation of .17 indicating very little consistency between the group and individual measures. The change in class B's performance on these subtests contributed strongly to the low correlation.

Considering both the data and the background information on classes A and B two factors need to be considered for future testing. One factor is setting. Class B's environment was primarily an individualized, small group instructional setting whereas class A's was primarily a large group seatwork situation. Therefore a group test situation was well within the experience of class A and would not have adversely affected their performance. Since class B was not as familiar with large group seatwork situations, the group testing may have adversely affected their performance. Since testing during Phase III-B will be conducted only with a small group of students (6 or less) this may minimize the setting effect.

The other factor is that of instructional materials. Despite the confusion on the addition with regrouping results, it does appear, that for content taught primarily with manipulatives, a paper-and-pencil test of that content may underestimate a student's knowledge. This underestimation may decrease over time -- i.e. may be less of a factor at the end of the school year after more opportunity for transference has been provided. However, even if that is the case we would still have a problem in that the growth scores (gains over time ) might be overestimated

if the pretest yielded an underestimate and the posttest did not. Therefore, it is important that we obtain information on classroom programs during Phase III-B so that we will be aware of where the effects of manipulative based programs may occur.

### STUDY 3

#### READING COMPREHENSION IN GRADES 2 AND 5

##### Rationale

The reading comprehension measure in the BTES group-administered test battery consists of short reading passages followed by multiple-choice questions. This is a standard format for a large scale testing operation, where measures must be group-administered and scoring of responses must be straightforward. The multiple-choice items are intended to assess the student's understanding of what he has read. This is, however, no simple matter. Reading comprehension is a complex partially subjective process. It is difficult to specify exactly what a reader should know when he has comprehended a passage. It is reasonable to expect the reader to understand facts directly stated in the text; but remembering facts is often the least important part of comprehension. The good reader organizes and synthesizes information; he relates what he is reading to past experience; he evaluates the importance of different parts of the text. In all these areas of higher-order comprehension, it is difficult to design questions with clear-cut right and wrong answers. In a multiple-choice question, one response needs to be clearly "best" while the distractors need to be somehow related to the story (in order to attract any responses) and yet in some way clearly wrong. Designing such questions is a challenging task. In some cases, questions may even discriminate against the thoughtful or imaginative reader, who can find plausible reasons for selecting supposedly "incorrect" alternatives.

In order to refine the reading comprehension items in the BTES battery, students were asked to respond individually to the multiple-choice questions

and to explain their responses. The goal was to locate items or answer choices that were confusing or directly misleading, especially for generally good readers. These items could then be eliminated or revised for future testing.

#### Description of Tasks

The group-test battery consists of several stories, each followed by a set of multiple-choice questions. About 45 minutes are allotted at each grade level for completion of the comprehension items. The group tests are included in Appendix C.

A subset of these items was selected for follow-up individual testing. Item analysis data were studied in order to select items that appeared to have potential problems. For instance, an item was included in individual testing if more students chose a particular distractor than chose the correct response, or if the students who chose a particular distractor had high scores on the total test. In other words, if a lot of students chose a particular response and/or if good readers chose a particular response, we wanted to know why. In Appendix C, the items included in individual testing have a circle around the item number.

In the individual testing session, each student was asked to read the stories, questions and answers out loud. For each question, the student was asked to select the best answer and explain why he made that choice. Follow-up questions were used to elicit more information about the student's reasoning processes with regard to particular response choices. For instance, the interviewer might ask "How do you know that ...?" or "What about answer 'B'?" If the student chose an incorrect answer, he was probed about the

"correct" alternative. Specific distractors were also designated for follow-up probes. Students were told in the beginning of the session that the interviewer wanted to understand what they were thinking about the questions and answers, so that we could improve the tests.

#### Design and Procedure

The group tests were administered to three classes at grade two and two classes at grade five. Standard procedures for BTES group testing were followed. The group testing session took 45 minutes at each grade level. Group testing for grade five took place in January, 1976; group testing for grade two took place in February.

Students were selected for individual testing on the basis of their group-test scores. An above-average score range was established, such that students would be able to deal reasonably successfully with the reading task but would still be making errors on the comprehension items. At grade two the range of scores was from 21 to 35 correct out of 39 items. At grade five, the range was from 18 to 33 correct out of 42 items. Within this range, students were randomly selected for participation in individual testing. At each grade level 20 students were selected for individual testing. One student in grade two could not complete the testing, so the sample was reduced to 19.

Individual testing was carried out in two separate sessions for each student. Each session was from 20 to 30 minutes long. All sessions were audio tape recorded. Grade five testing took place in early February; grade two testing took place in late February and early March, 1976.

Individual testing responses were coded on a rating form during the testing session, with the tape as a back-up. The student's initial answer

choice was coded as correct or incorrect. The student's response when asked to explain his answer was placed in one of three categories--plausible reason, unacceptable reason, or no reason. Sometimes, during the process of probing the student's reasons, a student would change his answer. The new choice was then recorded, and the reason for the new choice was evaluated according to the same three categories of plausible, unacceptable, or no reason.

The interview and scoring procedures were pilot tested with a small number of students. Some modifications were made at that time and procedures were standardized. Acceptable follow-up questions and conditions for their use were outlined, but the exact sequence of questions was expected to vary depending on student responses. No attempt was made to anticipate and precode all possible student reasons for responses. The interviewer's own understanding of the passage was expected to serve as the basic criterion for judgment. All interviewing and scoring were done by one person.

### Results

The major focus of attention in this analysis of reading comprehension was on the reasons students could give for their answers to multiple-choice questions. Among the items included in individual testing, there were considerable differences in the number of times students gave plausible reasons for "incorrect" answers. Some items appear to be good measures in their present form. In other cases, the interview responses helped pinpoint confusing or misleading items, which can be deleted or revised for Phase III-B testing. Representative item data will be discussed in this results section, but first some general background statistics will be presented.

General performance data. Students generally performed better in the individual testing situation than in group testing. One measure of performance in individual testing is the number of correct answers given as initial responses to test questions. This measure is directly parallel to the group test score, except that students read the passages and questions aloud and respond orally. Performance improved by 10 percent at both grade levels under individual testing conditions. Presumably, the student's greater attention to the task and more thorough reading of the passages account for this difference.

A second performance measure in individual testing is the number of correct answers including changes during the interview process. As the students explained their answers, they sometimes changed their minds. Students were told at the beginning of the session that the interviewer would not tell them which answer was right, that we wanted to know what they thought about the different parts of each question. If a student decided on a different alternative as he reasoned about the question, he was allowed to change his answer. Students changed right answers to wrong ones as well as wrong to right, but the overall balance was an increase in "correct" answers. The number of correct answers increased about seven percent as a result of changes during the interviewing. This was true at both grade levels.

Although the level of performance was generally higher during individual testing, the rank ordering of students remained relatively consistent from one situation to the other. The correlation between group-test scores and initial-response scores in individual testing was .80 at grade two and .75 at grade five. The group test gives a reasonable estimate of differences between students.

Item data, grade 2. On the whole, the second grade students selected a high proportion of correct responses. In individual testing, 77 percent of their initial choices were correct. A few items did cause problems. Three troublesome items are shown in Table 6.

These items are of interest because they were missed by a number of students, most of whom had some kind of plausible reason for the alternative they chose. In each case, the students eliminated at least one wholly unacceptable alternative and selected the response that would be judged a priori as "next best" to the correct choice. In effect these "next best" responses seem to be acceptable answers, to both the students and the interviewer.

In considering their reasons, it becomes clear that reasons were more often judged "plausible" when the interviewer herself was unsure as to why an answer was right or wrong.

In higher-order comprehension items, it is difficult to define "the" correct answer precisely. Items 3 and 7 are main idea questions, item 33 is an inferential relationship question. In each case, a rationale can be given for the pre-keyed correct answer. But justifications for answers at this level are largely complex rationalization. Psychologists are only now beginning to specify the factors involved. If a number of competent readers offer an alternative rationalization, their reasoning should be considered carefully. A test item that is so ambiguous should be revised or eliminated. A study of the reasoning used by students can contribute to our understanding of comprehension processes and to the defensibility of future theories.

In item 3, five students chose to title the story "The Monkey," and all stated that this was because Bess liked the monkey best. The implication

TABLE 6  
Confusing Grade Two Items

| Item                                               | Number of Students       |                    |
|----------------------------------------------------|--------------------------|--------------------|
|                                                    | Incorrect<br>with Reason | Total<br>Incorrect |
| 3. Which of these is the best title for the story? | 5                        | 7                  |
| *A. "The Circus"                                   |                          |                    |
| B. "The Elephant"                                  |                          |                    |
| C. "The Lion"                                      |                          |                    |
| +D. "The Monkey"                                   |                          |                    |
| 7. What is the main idea of this story?            | 5                        | 8                  |
| A. Indians had many horses.                        |                          |                    |
| *B. Indians used runners to send messages.         |                          |                    |
| +C. An Indian ran from village to village.         |                          |                    |
| 33. Why does the joey stay in the pocket?          | 8                        | 8                  |
| *A. To keep safe and warm.                         |                          |                    |
| B. To learn about pockets.                         |                          |                    |
| +C. To get milk to drink.                          |                          |                    |

\*This is the "correct" answer.

+This is the "incorrect" response which was plausibly defended.

is that a story title could refer to a key element of the story; on this basis the answer was judged plausible. Asking for a story title is a frequent way of trying to elicit the child's understanding of the main idea, but it is an ambiguous one. A story title doesn't necessarily express the main topic, so the student has to understand that this is what is really being asked for. Considering the confusion over this item, it is probably best to use a more direct question, such as "What is this story mainly about?" and eliminate all reference to story titles.

As item 7 illustrates, even more directly phrased main idea questions can be difficult. On item 7, five students chose response C. Their reasons centered around the fact that the major event of the story, what happened most, was that an Indian ran from village to village to tell people. These students have correctly grasped the main action; they seem to have difficulty with the abstract language used to generalize this action. They miss the word "messages" and ignore the fact that answer C says nothing about telling people, even though they include this in their verbal statements of the reason. They have trouble with the nominalization "runners." One student said that the story was "only a little bit about runners; they said a lot about the Indian running ..." Statements of main idea or topic often use abstract language and condense elements of the story; good readers learn to deal with these language elements. It may not be appropriate at this age level, however, to require abstraction from students who are successful at the basic task of sifting out major events.

The third grade two item to be discussed is item 33, an inferential relationship item. In individual testing, five students chose both A and C as equally acceptable answers. One student, when pressed to choose the

best answer, said "they're both the same but I picked A." Both A and C give reasons why the pocket is important to the baby; both require some inference. Alternative A was designated as the best answer because it is a broader, more inclusive response, which explains why the baby kangaroo stays in the pocket for so long. To select alternative A requires more reasoning independent of the passage. The subtlety of the distinction and the heavy reliance on extratext reasoning appear to make this item inappropriate for our grade two test.

Item data, grade 5. At grade 5, there were five items which presented problems. These items, and item data, are shown in Table 7. The items that caused problems at grade five are strikingly similar to those that were just discussed for grade two.

Item number 2 is an inferential relationship item similar to item number 33 at grade two. In both items, the plausible alternative is one where the student integrates information from consecutive sentences to produce a relevant but incomplete relationship. To select the best answer, the student must go beyond the text and reason about causal relationships. In this case, the grade five student might recognize that France being attacked is a general condition, while the implied and more specific motivator is that Alice didn't want to be caught. The extratext reasoning required in these items appears to be so judgmental and complex that the items are inappropriate for the test.

The next three grade five items, items numbered 13, 27, and 36, are all main idea questions, a class of questions that also caused difficulty at grade two. Analysis of the items points to the difficulty of defining the main topic in some passages and to the need for a clear question to elicit the main idea.

TABLE 7

Confusing Grade Five Items

| Item                                                       | Number of Students       |                    |
|------------------------------------------------------------|--------------------------|--------------------|
|                                                            | Incorrect<br>With Reason | Total<br>Incorrect |
| 2. Why did Alice Dubois leave France?                      | 11                       | 18                 |
| +A. France was attacked by its enemy.                      |                          |                    |
| B. She wanted to visit England.                            |                          |                    |
| C. She wanted to become a spy.                             |                          |                    |
| *D. She didn't want to be caught by the enemy.             |                          |                    |
| 13. What is the author's purpose in writing this story?    | 6                        | 9                  |
| *A. To compare gorillas and humans.                        |                          |                    |
| B. To describe the world's largest gorilla.                |                          |                    |
| +C. To announce the discoveries of two types of gorillas.  |                          |                    |
| D. To name the playful activities of gorillas.             |                          |                    |
| 27. Which sentence in the story best states the main idea? | 8                        | 18                 |
| +A. The first sentence.                                    |                          |                    |
| B. The second sentence.                                    |                          |                    |
| *C. The third sentence.                                    |                          |                    |
| D. The fourth sentence.                                    |                          |                    |
| E. The fifth sentence.                                     |                          |                    |
| 36. What is the author mainly doing in this story?         | 8                        | 12                 |
| +A. Telling about a funny incident.                        |                          |                    |
| B. Building up suspense.                                   |                          |                    |
| C. Describing where the boy lives.                         |                          |                    |
| *D. Introducing Kate.                                      |                          |                    |
| E. Setting an unhappy mood.                                |                          |                    |
| 40. Where does the main action of the story take place?    | 7                        | 11                 |
| +A. On a city sidewalk.                                    |                          |                    |
| *B. In front of the boy's house.                           |                          |                    |
| C. On a staircase.                                         |                          |                    |
| D. On a country road.                                      |                          |                    |
| E. In a small town.                                        |                          |                    |

\* This is the "correct" answer.

43

+ This is the "incorrect" response which was plausibly defended.

In item 13, the difficulty seems to be as much with the passage as with the students. The passage combines narrative with exposition; and the narrative of the class at the zoo seems mostly to obscure the structure of the passage. The first paragraph should state what the passage is about, and it doesn't. The student has to work throughout the passage to create some overall structure. One structure that students create is based on the comparison of the two kinds of gorillas, alternative C. This is one of the most clearly stated expository topics, and for these students it does represent the main idea. A student who integrates and processes further may discover that a more frequently occurring and broader theme is the comparison between gorillas and humans. But this theme is not stated very directly. It is unfair to penalize students for the poor quality of the prose. If this passage is to be used, it should be rewritten.

In item 27, the conflict is between sentence one and sentence three as a statement of the topic. It appears that these sentences are equivalent in the information they convey. (At the opposite extreme from the gorilla passage, this passage emphasizes the topic through redundancy in the first paragraph.) The addition of the somewhat more specific word "customs" in sentence three is offset by serial order and by the fact that the word "Etruscans" appear explicitly in sentence one and is referenced by a pronoun in sentence three. The best approach to improving this item would be to change the format to ask directly about different topics, where the choices are clearly differentiable.

In item 36, the difficulty seems to be with the judgment involved in deciding the author's purpose. More specifically, a number of students keyed on the fact that the story was funny. The selection is in fact a funny narrative, which at the same time introduces the main character. To

decide between these two purposes is largely a matter of judgment. It appears that the item would be improved by eliminating the response alternative which emphasizes humor.

The final item to be discussed is item number 40, an inferential description item. The students are drawn to choice A because the word "sidewalk" appears in the text and because a sidewalk is shown in the picture. To arrive at answer B, the student has to recognize that the boy runs out of his house into the street, so the action is in front of his house and occurs in the street as well as on the sidewalk. Because a large part of the action does take place on a sidewalk, and because of the complexity of reasoning in choice B, the item is a poor one. It should be eliminated from further testing.

#### Summary of Conclusions

At each grade level, a few items appeared to have two acceptable answers. Three second-grade items and five fifth-grade items fell in this category. These items should be revised or eliminated from future testing.

An analysis of these troublesome items indicated that a few general types of problems were occurring. These problems were similar across grade levels.

Some inference items required too much reasoning beyond the information in the text. In these items, the plausible alternative was at least partially correct and was more directly linked to language used in the story. In these items, students tended to see both answers as correct; and they should not be forced to choose between them.

Another type of question that caused problems was the main idea

question. Some main idea questions worked well, but others were less successful for a variety of reasons. It appears that main idea questions should be stated directly, rather than asking about story titles or the author's purpose. Especially at grade two, care should be taken not to require language that is too abstract. And sometimes the passage is at fault for not making passage structure or main ideas clear in the first place.

#### STUDY 4

#### TESTING FORMATS IN GRADE 5 MATHEMATICS

##### Rationale

Several of the items or scales within the Grade 5 mathematics test did not show much gain between the A and B testing dates. Some of these low gains can be attributed to a lack of opportunity to learn; others can be attributed to level of difficulty. Low gains do raise questions, however, about the appropriateness of some of the items. Grade 5 individual testing was primarily an exploratory effort to provide insight into sources affecting student performance. Due to the limitations of time and personnel, focus was narrowed to the areas of multiplication, word problems and geometry.

##### Description of Tasks

Group administered multiple-choice items from the BTES battery are shown in Appendix D. Those items selected for individual testing are identified in the Appendix.

Multiplication. The multiplication scale in the group-test battery (5M3) has a multiple-choice format. In real world situations, students would not have responses to select from but would simply work the problem. For this reason, free-response performance represents criterion for this study. Because multiplication problems are complex and involve many steps, it is difficult to design multiple-choice distractors that reflect all of the common answers students arrive at when working the problem. This discrepancy, as well as general problems of group testing, may influence the relationship between performance on a multiple-choice group test and performance on a free-response format of testing.

The 15 multiplication items in 5M3 were administered in both a multiple-choice and an open-ended format, in order to obtain an estimate of the correlation between the two formats. Since it is generally believed that the correlation between these two formats on computational tests is high, thus supporting the validity of the multiple-choice format, we looked to see whether that assumption could be confirmed in this situation. This double test administration also provided information on the diagnostic value of the distractors that supposedly reflect specific error patterns as well as information on how representative our distractors are in comparison to the distribution of the students' incorrect responses.

Geometry. Items on recognition of lines and figures (5M19 and 5M20) are being included because some of these items show lower correct response rates than might be expected. For example, only 49% of the students at the B (December) testing were correctly able to identify a rectangle imbedded in a larger figure. It's hard to believe that only half of the students in 5th grade know what a rectangle is. Certainly the items test more than simple recognition of a figure but it seems equally unlikely that the item complexity (format) is solely responsible for the low values. Through individual testing we should be able to obtain more accurate information on the student's knowledge of lines and figures as well as their ability to deal with specific item formats.

In individual testing, the students were first asked to draw the figure named by the interviewer. The following nine geometric lines and figures were dictated in the order listed here and represent the content of the 12 items in 5M19 and 5M20.

1. parallelogram
2. rectangle
3. perpendicular lines
4. parallel lines
5. square
6. right triangle
7. angle
8. line segment
9. line

After completing the dictation items the students were asked to read and solve the group test items which corresponded to those lines and figures that he or she was able to draw correctly. Responses to both sections were coded as right/wrong on the coding sheet.

The 5M16 scale (perimeter) in the group test battery includes 2 items using arbitrary units of measure, 3 items with standard units, and 3 items all of which refer to a single scale drawing. Students (per A & B test results) are more able to solve the items involving standard units. However, a number of students who are able to solve the standard items successfully are not able to solve the same type of problem when it refers to a scale drawing. Therefore, the 3 standard items and the 3 items referring to the scale drawing were included in the group test. Students who solved the standard perimeter problems correctly but were incorrect on the scale problems were selected for individual testing.

In individual testing, the student was asked to draw a rectangle or a box and was then questioned on the meaning of perimeter. Using the rectangle they were asked to show what they would be measuring if they were measuring the perimeter of that box or rectangle.

Regardless of the student's response to the question on the meaning of perimeter, each was asked to read and solve the first two items which referred to the scale drawing. Prior to working the items, the students read the information about the drawing and were questioned on what it

meant (why wasn't it drawn to the size it is supposed to be?, etc.).

For the last question, the students were referred back to their answer on perimeter and their answer was either confirmed or redefined for them. Responses were coded as right or wrong and whether scale conversion was applied. For the item on perimeter it was noted whether or not they counted edge segments, boxes, or computed their answer.

Word Problems. Eight word problems from the BTES battery were selected for group testing with individual follow-up. Five of these were items that showed considerable differences between the proportion of correct responses on the word problem versus the cross-referenced computation problem. For example, word problem #2 (see Appendix D) on the Danube River is a one-step word problem involving subtraction. On the 'B' test session 66% of the students solved the computation problem correctly but only 38% successfully solved the word problem. That sort of discrepancy suggests either that the students are not able to identify the operation required by the problem (a legitimate word problem skill) or that the reading level is too difficult for them (an undesirable source of difficulty). Some of the problems involve a two-step solution and these were selected as items for which information could be gained on the processes students use in analyzing the structure of such problems. Of particular interest is the ability to apply study skills in a logical step by step solution. Two other items (#6 & 8) were included because they appeared to be conceptually too difficult for the students. If in the individual testing the students were not able to understand the problem even with some guidance or prompting then the initial assumption would be confirmed and the items should be deleted.

In individual testing, students were first asked to read the problem

out loud. They were then questioned on how they would solve the problem. Depending on the response the student either computed his solution, was asked to explain the question in his own words, or was prompted to discuss certain sections within the word problem. If the student omitted pertinent information in his solution the interviewer would prompt him to recheck the information in the word problem. In most cases the interviewer worked with the child in an effort to help him or her reach a correct solution. In this way a better insight could be gained into the child's problem solving ability and his specific weaknesses within that process. Responses were coded for reading ability, trouble selecting the operation, use of appropriate information, operation finally selected, computational errors, inclusion of appropriate information with prompting, and additional notes on general ability and problems encountered. For data purposes, items were scored as right or wrong on the basis of the student's solution prior to prompting and/or discussion.

#### Procedure

The testing was conducted with a 5th grade class that had participated in the BTES study from A to B test dates. Due to the complexity of the classroom organizational structure (and changes in that structure) log information was not fully completed by the teacher and this class had to be dropped for our data purposes. This class tested somewhat lower than the other classes on some scales but was still within an appropriate range to be used for individual testing purposes. Twenty-five students completed all group-administered tests. Fifteen of these students were tested individually. Selection was based primarily on their scores on the geometry items.

Testing was carried out in three phases: (1) The group-administered multiple-choice items were administered to all students in the class. Thirty minutes were allocated for this test session; all but three students finished within this time limit. Following normal BTES procedure, students were instructed to skip an item if they did not understand it and if their answer would simply be a guess. (2) Following the group test, the fifteen individual interviews were conducted. All individual sessions were audio-taped. (3) A second group-test session was held to administer the open-ended multiplication items. This session took place eight days after the initial multiple-choice test. All students finished the test in about 10 minutes.

#### Results and Discussion

Multiplication. The raw mean pretest score (uncorrected) was 9.08 (15 items) with a standard deviation of 4.02. The posttest mean was 7.84 with a standard deviation of 4.54. The Pearson product moment correlation was .62. Using corrected scores (-.333 for each incorrect response) the pretest mean fell to 7.33 with a standard deviation of 5.35. With these corrected pretest scores the pre and posttest correlation remained the same at .62.

This mid range correlation seemed to be partially attributable to test conditions on the pretest. The classroom layout is such that students do not have separate desks but rather sit at small learning center tables which accomodate 3-4 students. The opportunity for "copying" another student's answers on the multiple choice pretest was high and a few students appeared to take advantage of that opportunity. Three students in particular scored in the middle to high range on the pretest and were selected for individual testing. Though individual testing did not include the

multiplication problems, it was evident from the word problem section that these three children did not know their basic facts. The posttest scores for these students dropped 7 to 9 points. It appears that the probable incidence of "cheating," even with 3 out of 26 students, can have highly adverse effects on the data and we may need to tighten our testing procedures. In classrooms that do not offer the opportunity to sufficiently separate the children we may consider providing portable cardboard separations, using a more appropriate room for testing (if possible), or testing with smaller groups. Certainly the fluctuations in students' moods and motivation, stability of knowledge, external disruptive influences, etc., contribute enough error variance that the situation should not be further complicated by test conditions that are more amenable to our control.

Responses to the open-ended items gave useful information about the frequency of different types of errors. Based on the posttest results, 40% of the students could be considered proficient in multiplication. Their work reflected a good understanding of the multiplication algorithm marred only by careless errors. The errors were primarily in basic facts or failing to add what had been carried from a previous column. Their average rate of carelessness was 14%. About 16% of the students were partially proficient in the multiplication algorithm. Their work reflected systematic error patterns with problems involving 2-digit multipliers. Most of these error patterns were compound ones involving a combination of two systematic errors or single error patterns compounded by a wide variety of careless errors. An example of a double error pattern was shown by one student who failed to line up the partial products correctly plus consistently failed to add what had been carried for the

second partial product. Needless to say, such complex error patterns would be extremely difficult to detect on a multiple choice exam offering three distractor choices per item. However, one student's pattern was detected on the pretest and confirmed by the posttest. She had a single systematic error of failing to line up the partial products correctly. She got 9-items wrong on the pretest and the same 9 items wrong on the posttest (all of the 9 items involving 2-digit multipliers). On 5 out of 9 items her posttest answers matched her pretest distractor choice. On the other four problems the partial product error pattern was compounded by additional computational errors.

If these results can be generalized to a larger school population then we may conclude that only 4% of the students have single systematic errors (in multiplication) that could be detected by a test such as ours. Considering the amount of time required to analyze the individual answer choice patterns for a large test population it would seem impractical to invest much energy in trying to diagnose the error patterns for such a small number of students.

This is not to say that a careful selection of distractors is meaningless. Certainly we would want the distractors to reflect common errors children make and thus provide for more useful items. In analyzing the posttest answers, 31% of the pretest distractors were represented. The distribution of incorrect responses did provide some useful information on possible alternate distractors that could be incorporated in the final test battery. Examples would be answers resulting from the students treating a zero as a one (Example A) and the one's digit instead of the ten's digit being carried (B).

$$\begin{array}{r} \text{A.} \quad 278 \\ \quad \times 30 \\ \hline \quad 278 \\ 834 \\ \hline 8616 \end{array}$$

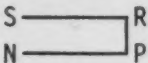
$$\begin{array}{r} \text{B.} \quad 37 \\ \quad \times 6 \\ \hline 204 \end{array}$$

The incorrect item responses which were repeated by 4 or more students are candidates for future item distractors. However, only 4 out of the 25 items showed that many incorrect response repeats. The most difficult test item was solved incorrectly by 19 students and each of the 19 responses were different. When answers become that unique it is improbable that discriminating distractors can be generated.

Geometry (lines and figures). On the lines and figures items, the primary concern had been that item formats using letter notation might be confusing students and thus reducing scores. Individual testing supported this hypothesis. Table 8 shows student performance on the multiple-choice group test and on the individually-administered production task.

TABLE 8  
Performance on Lines and Figures Items

| Figure              | Proportion Correct |                       |
|---------------------|--------------------|-----------------------|
|                     | Group Recognition  | Individual Production |
| Square              | .73                | .93                   |
| Rectangle           | .67                | .87                   |
| Parallel lines      | .60                | .33                   |
| Line segment        | .53                | .27                   |
| Parallelogram       | .67                | .20                   |
| Right triangle      | .13                | .13                   |
| Angle               | .33                | .07                   |
| Line                | .40                | .07                   |
| Perpendicular lines | .00                | .00                   |

Notice that for the square and rectangle items performance was better on the production task. Almost all students know these two basic figures. But students are relatively unfamiliar with geometric notation, as became apparent during individual administration of group-test items. In approaching the test item with the imbedded rectangle some of the children viewed each part of the figure as separate and were confused by the extra lines. Within the rectangle SRPN there is the triangle SON. Since the letter O was within the rectangle some students thought that it must be accounted for in their answer. One student thought that the answer choices were words and that the object of the question was to find the word that was in the picture! Another student had difficulty since he did not realize that the starting point is also assumed to be the ending point. Therefore in tracing SRNP his figure was incomplete (e.g. ). As time progressed the interviewer started instructing the children to trace the figure prior to answering the question and this of course biased the results. None of the students who were asked to find the rectangle, trace over it, and then determine which letters named it had any problem with the question. Also the order in which the letters were read did not appear to present any difficulties. Those students who were initially incorrect were able to solve the problem when asked to follow the above procedure. It therefore appears that items that emphasize geometric notation introduce a test-taking factor that is separate from knowledge of geometric figures and is probably inappropriate for fifth graders. Items should be revised for future BTES testing to eliminate errors from geometric notation.

Individual testing suggested a more appropriate way to make the test item on the square more difficult. In asking the children what the

difference was between a rectangle and a square the most frequent initial response was that a rectangle is longer and thinner than a square. When asked what the relationship was between the sides (interviewer assigned numerical values to two opposing sides of the rectangle and to one or two sides of the square) only 2 out of 10 children knew that the sides of a square are equal to one another. However, 4 out of 5 knew that the opposite sides of a rectangle are equal. This suggests that items which go beyond obvious visual difference and require careful consideration of measurements will be more difficult and potentially more discriminating.

Table 8 also shows the difficulty of the production tasks other than square and rectangle. Students beginning to learn lines and less familiar figures do better on the group-administered recognition test. Straightforward recognition items such as these, without confusing notation, seem to be in the right difficulty range for fifth graders and are likely candidates for good test items.

Geometry (perimeter). The issue of interest in the perimeter items is the discrepancy in performance between standard unit items and a scale drawing application item. Individual testing provided insights into the sources of difficulty in the scale drawing problem.

Of the ten students who had answered all three standard perimeter items correctly on the pretest, six were able to define perimeter correctly in the individual testing. Two were incorrect, one was borderline, and the last one had no response, as did the other four students who were questioned.

When the children were asked to read the information which referred to the scale drawing, with few exceptions they ignored the sentence about scale conversion and had to be specifically prompted before they

read it. When asked why the pattern was drawn to scale most stated that the figure would be too big for the page if it were drawn to actual size.

In responding to the first item on width, most of the students knew the difference between length and width and did not appear to refer to the labels provided. If a child was unable to identify which side was the width, she was prompted to look at the figure to see if that information was given. Surprisingly it took a couple of children a minute or so to locate the label. It almost appeared as if they had problems focusing on and identifying the different points labeled within the diagram.

For the first two items on length and width, six students were able to solve them correctly without prompting (of course this was after the diagram discussion). Five of these students correctly explained their solution (counted by two's or multiplied the number of boxes by two). Of the remaining eight students, two were able to do the problems correctly with a little prompting and the others failed to apply scale conversion even with prompting. These students did not seem to recognize that the scale information was an integral part of the problem and needed step-by-step help in order to solve the problem.

On the final item, two out of the six students who had previously defined perimeter correctly were able to solve this problem without prompting. Both of them computed their answer and used their answers to the previous items on length and width in that computation. Of the remaining four, two counted the boxes and naturally were incorrect (see conclusions), two could not apply their previous definitions of perimeter but with prompting solved it computationally without applying conversion. Of the four other students who were given the definition of

perimeter and asked to solve the problem, two did so computationally but forgot to apply scale conversion and the other two needed step-by-step help.

Of the five students who had answered this last item correctly on the pretest, three of them were among the six that defined perimeter correctly and two were able to solve the problem correctly in the individual testing.

Unless the reactivity data shows otherwise, these three items appear to be inappropriate for a number of reasons. They seem to require a fairly high level of reading comprehension skills in that the students must be able to scan the material to locate needed information. If students do not fully scan the material they may fail to notice the conversion information and proceed to solve the items without realizing that conversion is required. If they do recognize that conversion is necessary, they must retain that knowledge and apply it to each step. Here the problem of multi-step items becomes more evident. A child may apply conversion in computing the length and width but unless he uses his answers to these two items in computing the perimeter he is quite likely to forget about conversion in his effort to figure out the perimeter. The children appear to be approaching each item separately and are more likely to return to the scale drawing to "compute" perimeter than to use their answers on length and width. By returning to the drawing the presence of arbitrary units adds another factor of difficulty. In solving length and width the student may count either the edges or the boxes and arrive at the same answer. However, if counting is used to solve perimeter then the edges and not the boxes must be counted (unless boxes for each side are counted separately). Most students are not sophisticated enough

to realize that by counting the boxes all the way around their answer will be reduced by 4 units (8 with conversion) in comparison to counting the edge segments.

These factors of reading comprehension, multi-step solutions, and use of arbitrary units are probably the primary contributors to the low proportion of students passing this item among those who were able to solve the standard perimeter items correctly.

For test purposes it would, probably be best to limit ourselves to the standard perimeter items with a little refinement. Since four out of the ten students who solved the three standard perimeter items correctly were not able to define the term perimeter, one suspects that these items are not as sensitive as they might be. The inclusion of nonessential data (e.g. height of the polygon) in one or more of the items would probably discriminate between the students who understand perimeter and those who are merely showing a logical approach in the manipulation of the data provided. We may also consider including items which require the student to select the best definition for perimeter versus area and volume.

Word Problems. Eight word problems were included in the group multiple-choice test, and these same eight were used in the individual testing. This particular class had spent very little time on word problems and this was reflected in their lack of word problem attack skills. The children had no difficulty reading the word problems, which reduced concern about reading level. The children had considerable difficulty comprehending the problems. They frequently ignored essential information. (Ten out of fifteen students ignored the word "each" in problem number one. Twelve of fourteen ignored the phrase "less than Anne" in problem number 4.) For problems numbers 1, 2, 4, and 5 errors were primarily related to poor

study skills. When asked to reread the problem or review a specific statement most students easily recognized the data or step that they had omitted. Such factors make the problems difficult, but in a way that seems valid to the skill being tested. An examination of reactivity data will indicate whether students who receive instruction in word problems are learning better study skills. Hopefully, this will be the case. If not, it may be necessary to select the simpler items for future testing.

Problems number 3 and number 7 were more difficult conceptually. Although the students had the computational skills necessary to solve the problems they had difficulty recognizing the appropriate operation. Step two of problem number 3 was most often solved by the use of an array diagram rather than by division or the use of factors. Instead of multiplication repetitive addition was used by most of the students in solving problem number 7.

Problem number 6, which required the computing of an average, was dropped during individual testing. The students stated that they did not remember learning about averages and the teacher confirmed that he had not covered averages. Problem number 8 was conceptually and computationally too difficult for these students. Only one student recognized the need to divide but he was unable to complete the division correctly. The need for division was explained to five other students and only one was able to complete the division without help.

Individual testing revealed a number of instances in which students were making computational errors. On problem number two, eleven students verbally identified subtraction as the required operation; only six of the eleven were able to carry out the computation correctly. Computational

skill is an extraneous factor in the measurement of word problem solving, so it seems appropriate to deemphasize the computational aspects of the word problems and increase the conceptual emphasis. Instead of asking the students to compute the answer they could be asked to recognize the manner in which the problem should be solved. Specifically they would need to recognize the correct computational construction necessary to solve the problem. The same solution could also be offered more than once with answers given and then the students would be required to verify the computation in order to select the correct answer.

#### Summary of Conclusions

The correlation between multiple-choice and open-ended multiplication was .62. A primary factor reducing this correlation was apparent copying of answers on the multiple-choice test. Small group testing in Phase III-B should increase our control over testing conditions and decrease cheating as a factor of variance. Responses on the open-ended multiplication items showed that most error patterns were compound ones that could not be diagnosed from the multiple-choice response patterns.

In geometry, students were unfamiliar with many of the line items and did better on the group recognition items than on the individual recall items. Recognition items on figures also appeared to be good candidates for test items, as long as care is taken to eliminate complex embedded figures with confusing geometric notation.

The scale drawing items on perimeter appear to test complex visual scanning and comprehension processes rather than direct application of perimeter. Other ways of increasing difficulty on standard perimeter items would be more appropriate for future testing.

Two of the word problems were conceptually inappropriate. The remaining problems were conceptually appropriate but responses indicated that some students were approaching the problem correctly but made computational errors. Reducing the computational level would contribute to a more sensitive group measure of problem solving skills.

REFERENCES

Filby, N. and Dishaw, M. Development and refinement of reading and mathematics tests for grades 2 and 5. (BTES Technical Report III-1). San Francisco: Far West Laboratory for Educational Research and Development, 1975.

Filby, N. and Dishaw M. Refinement of reading and mathematics tests through an analysis of reactivity. (BTES Technical Report III-6). San Francisco: Far West Laboratory for Educational Research and Development, 1976.

**APPENDIX A**

**Individual Testing Materials for Grade 2  
Reading Decoding**

**(Study 1)**

Task 1  
SIGHT WORDS

1. right
2. of
3. there
4. was
5. from
6. one
7. what
8. some
9. night
10. whose
11. know
12. where
13. have

Task 2  
WORD LIST

1. nest
2. chair
3. thin
4. pencil
5. kite
6. spoon
7. left
8. coat
9. pail
10. shirt
11. plate
12. beard
13. clown
14. hammer
15. mouth
16. palace
17. candle
18. turtle
19. truck
20. swing
21. whistle
22. flowers

Task 3  
WORDS IN CONTEXT

1. The bird made a nest in a tree.
2. Sally sat quietly in her chair.
3. Mary is fat but Jill is thin.
4. You write with a pencil.
5. The children will fly a kite in the park.
6. You eat soup with a spoon.
7. Robert left his paper at home.
8. Ann put on her hat and coat to go outside.
9. Sonia put some sand in her pail.
10. Jim put on his pants and shirt.
11. Pat ate all the food on her plate.
12. The man has long hair and a long black beard.
13. At the circus we saw a funny clown.
14. I hit the nail with a hammer.
15. Mark put the candy in his mouth and ate it.
16. The king lives in a palace.
17. Jan lit the tall candle.
18. An animal that moves very slowly is a turtle.
19. The man drove the truck down the road.
20. The swing on the playground is broken.
21. Peter blows a whistle to call his dog.
22. Jan picked a bunch of pretty spring flowers.

Task 4  
SHORT AND LONG VOWELS  
NONSENSE WORDS

- |          |          |
|----------|----------|
| 1. kip   | 14. nad  |
| 2. fube  | 15. jod  |
| 3. dap   | 16. seef |
| 4. toak  | 17. kipe |
| 5. lebe  | 18. dape |
| 6. fon   | 19. bim  |
| 7. jode  | 20. voal |
| 8. fain  | 21. kete |
| 9. ket   | 22. fub  |
| 10. vune | 23. leb  |
| 11. nade | 24. gaid |
| 12. meep | 25. fone |
| 13. bime | 26. vun  |

**APPENDIX B**

**Group and Individual Testing Materials  
for Grade 2 Mathematics  
(Study 2)**

**20**

Spring 1976  
Individualized  
Testing  
Group Pretest

NAME:

(Last)

(First)

**IA**

Directions: Circle the letter of the answer you choose.

SAMPLE:

$$3 + 1 = \square$$

A. 2

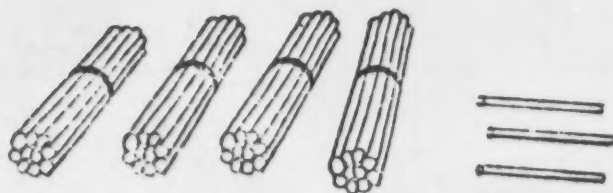
C. 4

B. 3

D. 5

5

How many TENS are there in this picture?



A. 3

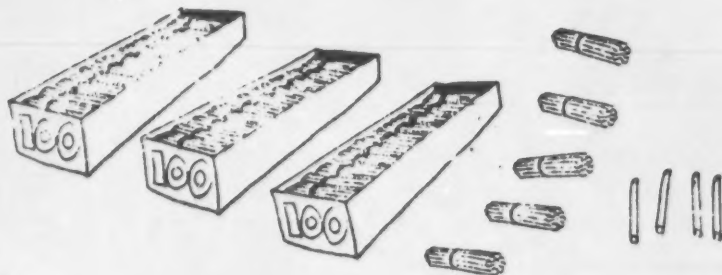
C. 0

B. 4

D. 7

6

How many HUNDREDS are there in this picture?



A. 3

C. 4

B. 5

D. 0

7

Which number is in the ONES place?

63

A. 6

C. 0

B. 9

D. 3

8

Which number is in the HUNDREDS place?

524

A. 2

C. 4

B. 5

D. 52

9

Which number is in the TENS place?

145

A. 1

C. 4

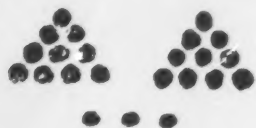
B. 5

D. 45

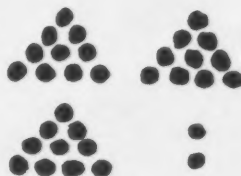
10

Which picture shows 23?

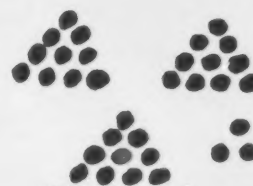
A.



B.

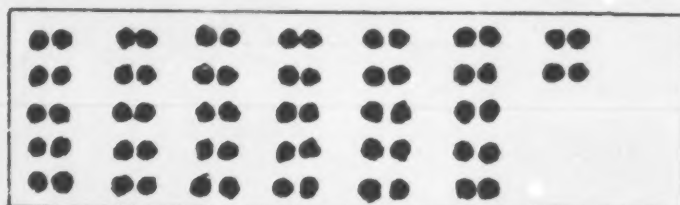


C.



11

How many dots are in this picture?



A. 7

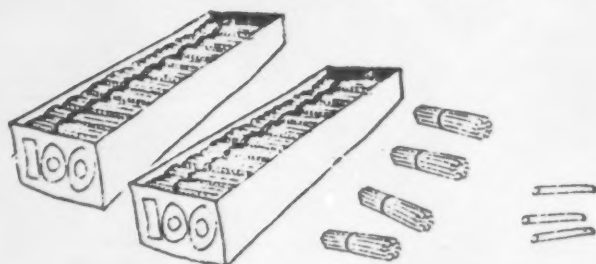
C. 74

B. 64

D. 46

12

Which number matches this picture?



A. 207

C. 603

B. 243

D. 2,043

13

Which number means 3 TENS and 4 ONES?

A. 7

B. 304

C. 34

D. 43

14

What is another name for 16?

A. 6 TENS and 1 ONE

C. 1 HUNDRED AND 6 TENS

B. 1 TEN and 6 ONES

D. 7 ONES

15

Another way to write 59 is:

A.  $5 + 9$

C.  $9 + 5$

B.  $90 + 5$

D.  $50 + 9$

16

Another way to write 975 is:

A.  $900 + 70 + 5$

C.  $900 + 7 + 5$

B.  $90 + 70 + 50$

D.  $90 + 75$

17

Which of these numerals show 4 HUNDRED AND 5 ONES?

A. 45

C. 4005

B. 405

D. 450

18

Another way to write  $300 + 60 + 8$  is:

A. 368

C. 30608

B. 300608

D. 3068

19

In the number 867, the number 6 means:

A. 6 ONES

C. 6 TENS

B. 6 HUNDREDS

D. 6 THOUSANDS

WHAT SIGN GOES IN THE  $\bigcirc$  ?

A.

245

 $\bigcirc$ 

145

A. = B. &lt; C. &gt;

B.

2 &lt; 2

 $\bigcirc$ 

1

A. = B. + C. -

6.

$1 + 7 \bigcirc 7$

A. &lt;

B. &gt;

C. =

11.

$3 \bigcirc 5 > 12$

A. +

B. -

C. =

7.

$6 \bigcirc 3 + 5$

A. &lt;

B. &gt;

C. =

12.

$9 = 16 \bigcirc 7$

A. +

B. -

C. =

8.

$6 - 4 \bigcirc 3 + 2$

A. &lt;

B. &gt;

C. =

13.

$5 < 3 \bigcirc 5$

A. +

B. -

C. =

9.

$853 \bigcirc 803$

A. &lt;

B. &gt;

C. =

14.

$70 + 10 \bigcirc 60 + 20$

A. +

B. -

C. =

10.

$3 + 7 \bigcirc 5 + 5$

A. &lt;

B. &gt;

C. =

15.

$8 - 2 = 7 \bigcirc 1$

A. +

B. -

C. =

16

$$\begin{array}{r} 2 \\ +3 \\ \hline \end{array}$$

A. 1

C. 6

B. 5

D. 7

21

$$\begin{array}{r} 9 \\ +4 \\ \hline \end{array}$$

A. 5

C. 14

B. 15

D. 13

17

$$\square - 4 = 4$$

A. 6

C. 9

B. 7

D. 8

22

$$\square = 7 + 7$$

A. 13

C. 15

B. 14

D. 16

18

$$1 = \square - 6$$

A. 5

C. 7

B. 9

D. 8

23

$$9 + 6 = \square$$

A. 15

C. 18

B. 16

D. 19

19

$$\begin{array}{r} 5 \\ +4 \\ \hline \end{array}$$

A. 9

C. 5

B. 7

D. 6

24

$$\begin{array}{r} 13 \\ +6 \\ \hline \end{array}$$

A. 15

C. 18

B. 17

D. 19

20

$$\square - 5 = 7$$

A. 14

C. 12

B. 13

D. 11

25

$$44 = \square - 3$$

A. 41

C. 47

B. 45

D. 48

# JUST FOR FUN

B-8

Connect the dots. Start with 50.



1

Which of these numerals  
names the smallest number?

- A. 390                      C. 397  
B. 379                      D. 400

3

Which number comes between  
55 and 62?

- A. 61                      C. 49  
B. 64                      D. 53

2

What number comes  
right after 889?

- A. 899                      C. 890  
B. 880                      D. 900

4

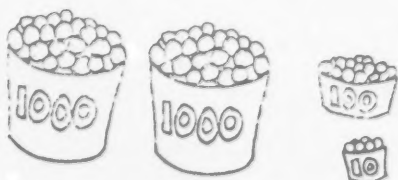
Which of these numerals  
names the largest number?

- A. 699                      C. 709  
B. 710                      D. 712

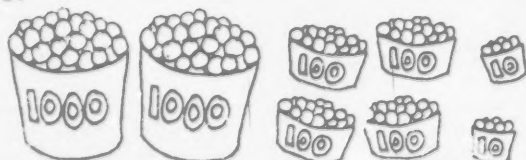
5

Which picture shows the most marbles?

A.



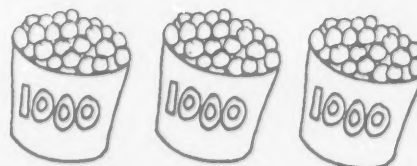
B.



C.



D.



20

$22 + 7 = \square$

A. 25

C. 28

B. 29

D. 31

25

$$\begin{array}{r} 36 \\ +42 \\ \hline \end{array}$$

A. 78

C. 14

B. 64

D. 88

21

$\square = 11 + 5$

A. 16

C. 13

B. 17

D. 14

26

$\square = 29 + 4$

A. 34

C. 23

B. 25

D. 33

22

$$\begin{array}{r} 10 \\ +55 \\ \hline \end{array}$$

A. 45

C. 65

B. 55

D. 60

27

$$\begin{array}{r} 37 \\ +6 \\ \hline \end{array}$$

A. 43

C. 33

B. 31

D. 44

23

$$\begin{array}{r} 72 \\ +22 \\ \hline \end{array}$$

A. 50

C. 90

B. 54

D. 94

28

$15 + 7 = \square$

A. 8

C. 22

B. 12

D. 32

24

$$\begin{array}{r} 23 \\ +34 \\ \hline \end{array}$$

A. 58

C. 47

B. 57

D. 11

29

$$\begin{array}{r} 58 \\ +8 \\ \hline \end{array}$$

A. 66

C. 56

B. 50

D. 68

30

$$\begin{array}{r} 49 \\ +32 \\ \hline \end{array}$$

A. 71      C. 77

B. 87      D. 81

35

$$\begin{array}{r} 57 \\ +13 \\ \hline \end{array}$$

A. 64      C. 70

B. 60      D. 44

31

$$\begin{array}{r} 45 \\ +28 \\ \hline \end{array}$$

A. 77      C. 63

B. 73      D. 67

36

$$\begin{array}{r} 173 \\ +105 \\ \hline \end{array}$$

A. 172      C. 288

B. 272      D. 278

32

$$\begin{array}{r} 345 \\ +122 \\ \hline \end{array}$$

A. 463      C. 467

B. 567      D. 267

37

$$\begin{array}{r} 30 \\ 18 \\ +42 \\ \hline \end{array}$$

A. 100      C. 80

B. 90      D. 86

33

$$\begin{array}{r} 23 \\ 12 \\ + 7 \\ \hline \end{array}$$

A. 42      C. 52

B. 32      D. 41

38

$$\begin{array}{r} 224 \\ +323 \\ \hline \end{array}$$

A. 545      C. 501

B. 547      D. 541

34

$$\begin{array}{r} 29 \\ +16 \\ \hline \end{array}$$

A. 45      C. 33

B. 43      D. 35

39

$$\begin{array}{r} 466 \\ +231 \\ \hline \end{array}$$

A. 697      C. 236

B. 695      D. 687

NAME: \_\_\_\_\_ (Last) \_\_\_\_\_ (First)

IB

1

Which of these numerals  
names the smallest number?

- A. 390                      C. 397  
B. 379                      D. 400

3

Which number comes between  
55 and 62?

- A. 61                      C. 49  
B. 64                      D. 53

2

What number comes  
right after 889?

- A. 899                      C. 890  
B. 880                      D. 900

4

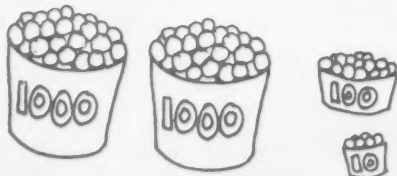
Which of these numerals  
names the largest number?

- A. 699                      C. 709  
B. 710                      D. 712

5

Which picture shows the most marbles?

A.



B.



C.



D.



WHAT SIGN GOES IN THE  $\bigcirc$  ?

B-14

$$245 \bigcirc 145$$

A. =    B. <    C. >

$$2 < 2 \bigcirc 1$$

A. =    B. +    C. -

6)  $6 - 4 \bigcirc 3 + 2$

A. <    B. >    C. =

7)  $3 + 7 \bigcirc 5 + 5$

A. <    B. >    C. =

8)  $8 \bigcirc 5 > 12$

A. +    B. -    C. =

$$9) \quad 1 = \square - 6$$

$$10) \quad \square - 5 = 7$$

$$11) \quad 44 = \square - 3$$

$$\begin{array}{r} \square \\ - 3 \\ \hline 44 \end{array}$$

1 2)

1 3)

1 4)

1 5)

16)

$$\begin{array}{r} 58 \\ + 8 \\ \hline \end{array}$$

17)

$$\begin{array}{r} 49 \\ + 32 \\ \hline \end{array}$$

18)

$$\begin{array}{r} 23 \\ 12 \\ + 7 \\ \hline \end{array}$$

Date \_\_\_\_\_

Name \_\_\_\_\_

B-18

| R   | GE | W/c E/R | W -IE | Ordered Correct   | Comments                                          |
|-----|----|---------|-------|-------------------|---------------------------------------------------|
| 1.  |    |         |       | Y N               |                                                   |
| 2.  |    |         |       | Y N               |                                                   |
| 3.  |    |         |       | Y N               |                                                   |
| 4.  |    |         |       | Y N               |                                                   |
| 5.  |    |         |       | Y N               |                                                   |
| 6.  |    |         |       |                   | Defined < , > Correctly <sup>Y</sup> <sub>N</sub> |
| 7.  |    |         |       |                   |                                                   |
| 8.  |    |         |       |                   |                                                   |
| 9.  |    |         |       |                   |                                                   |
| 10. |    |         |       | Reversed Rdg. Y N |                                                   |
| 11. |    |         |       | Reversed Rdg. Y N |                                                   |

| R-Writtn | R c M. | W-Writtn | W-M |                                     |
|----------|--------|----------|-----|-------------------------------------|
| 12.      |        |          |     | Used 1's only Switched from 1's Y N |
| 13.      |        |          |     |                                     |
| 14.      |        |          |     |                                     |
| 15.      |        |          |     | G.E. for dropping zero's Y N        |
| 16.      |        |          |     |                                     |
| 17.      |        |          |     |                                     |
| 18.      |        |          |     |                                     |

Exchanged 10 one's for 1 ten Y N

Knew the 1 being carried = 10 Y N

APPENDIX C  
Group Testing Materials for Reading Comprehension<sup>1</sup>  
(Study 3)

<sup>1</sup> Selected items from these tests were used for follow-up individual testing. These items are identified by a circle around the item number.

Grade 2

## READING COMPREHENSION

NAME: \_\_\_\_\_ (LAST) \_\_\_\_\_ (FIRST)

Once there was a little girl named Bess. She visited her grandfather who lived near a circus. There were many elephants, monkeys, and lions in the circus. Bess liked the monkeys best.

1. Bess was \_\_\_\_\_.

- A. large
- B. little
- C. old
- D. sad

2. Which animals did Bess like best?

- A. camels
- B. elephants
- C. lions
- D. monkeys

3. Which of these is the best title for the story?

- A. "The Circus"
- B. "The Elephant"
- C. "The Lion"
- D. "The Monkey"

Susan liked to do dangerous thing. She walked on top of a high brick wall. She climbed to the very top of the oak tree. She rode her bicycle very fast. One day she zoomed right into the street. A car almost hit her. Susan saw the car just in time.

- ④. Where do you think Susan would like to go?
- A. To the library
  - B. To a movie
  - C. To the park
- ⑤. What would Susan like to have for her birthday?
- A. A new dress
  - B. A toy truck
  - C. A pair of roller skates
6. What lesson could Susan learn?
- A. Be careful when you ride your bike.
  - B. Don't leave your bike out in the rain.
  - C. Always walk on the sidewalk.

An Indian put his ear to the ground. He heard many horses. They were coming his way. He ran to tell his people. Then he ran to tell the people of the next village. He was a runner. Indians used runners as one way to send messages.

7. What is the main idea of this story?
- A. Indians had many horses.
  - B. Indians used runners to send messages
  - C. An Indian ran from village to village.
8. Who did the Indian tell the message to first.
- A. The people at the next village
  - B. The people who owned the horses
  - C. The people at his own village
9. The Indian \_\_\_\_\_ from village to village.
- A. walked
  - B. ran
  - C. rode a horse

GO ON TO THE NEXT PAGE

Have you ever had a real adventure? George and I did. We met a hungry whale!

It began when George and I were out in the boat. Suddenly, George said, "Look! What's that?"

"A whale!" I said.

Sure enough. The whale came swimming up. He came right to the boat. Then he said in a voice like thunder, "I am a whale with a very big hunger."

George was brave. He opened our lunch and leaned over the side. "Ham, jam, or peanut butter?"

The whale ate them all. "That's all?" he said. And we could tell he was mad, for he dove deep down in the sea.

And that's where he saw our boat's rope. Before we could stop him he ate it.

Then he swam away. And as he swam, the rope pulled tight. One end was tied to our boat, one end was deep in his insides.

10. Which of these happened first?

- A. The whale swam away.
- B. The whale ate some pie.
- C. The boys opened their lunch.
- D. The whale came up to the boat.

11. What happened to the rope?

- A. The rope came off the boat.
- B. The boys lost the rope.
- C. The rope was tied to the whale's tail.
- D. The whale ate the rope.



12. The story says that the whale spoke in a "voice like thunder." What does that mean?
- A. The whale had a very loud voice.
  - B. The whale said that a storm was coming.
  - C. The whale talked fast, and the boys couldn't understand him.
  - D. The whale opened his mouth, and a thunderstorm began.
13. What made the whale dive deep down into the sea?
- A. He was mad.
  - B. He had too much to eat.
  - C. He wanted to hide from the boys.
  - D. He was going to meet another whale.
14. At the end of the story, what happened to the boat?
- A. It pulled the whale.
  - B. It carried the whale.
  - C. It was on top of the whale.
  - D. It was being pulled by the whale.
15. What is this story mostly about?
- A. A fishing trip
  - B. A ride in a boat
  - C. A whale who wanted to eat
  - D. Getting away from a whale

STOP

Here is a story Dick wrote:

Mary read two books last week. She can write well too. I wish I could do as well as Mary. When I get to her grade, maybe I can.

- (25.) Mary read two books last week. How many books do you think Dick read?
- A. more than two
  - B. less than two
  - C. two
26. How old do you think Dick is?
- A. not as old as Mary
  - B. older than Mary
  - C. the same age as Mary
- (27.) Dick shows that he \_\_\_\_\_.
- A. does not like to read
  - B. wants to read better
  - C. can read better than Mary

Read the story. Answer the questions.

Do you know about young kangaroos? All baby kangaroos are called joeys. Baby kangaroos live in their mother's "pocket." The mother kangaroo carries her joey with her, in her pocket.

A baby kangaroo is only one inch long when it is born. It has no fur. It cannot see. As soon as it is born, it crawls into the pocket.

The baby kangaroo stays in the pocket for four months. It drinks milk, sleeps, and grows bigger and stronger.



|                                                                                                                                      |                                                                                                                                                                   |
|--------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p>28] What is this story mainly about?</p> <p>A. baby kangaroos</p> <p>B. pockets</p> <p>C. a kangaroo mother</p>                   | <p>31] The baby kangaroo's eyes are _____.</p> <p>A. open</p> <p>B. closed</p> <p>C. blue</p>                                                                     |
| <p>29] What are baby kangaroos called?</p> <p>A. kids</p> <p>B. babies</p> <p>C. joeys</p>                                           | <p>32] When does the baby kangaroo climb into the pocket?</p> <p>A. right after it is born</p> <p>B. when it is four months old</p> <p>C. when it gets sleepy</p> |
| <p>30] Where do baby kangaroos live?</p> <p>A. in the mother's pocket</p> <p>B. in a soft nest</p> <p>C. in a hole in the ground</p> | <p>33] Why does the joey stay in the pocket?</p> <p>A. to keep safe and warm</p> <p>B. to learn about pockets</p> <p>C. to get milk to drink</p>                  |

Read the story. Answer the questions.

The big store looked very far away. Benny and Al wondered if they would ever get there. It was very hot and crowded on the sidewalks. They had to squeeze between the people as they ran. At last they came to the big glass doors. The doors swung open and they were soon inside the cool building. They rushed down the stairs to the toy shop. Right in front of them were the shiny skates they had come to buy.

|                                                                                                                                                  |                                                                                                                                 |
|--------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------|
| <p>34 What is this story mostly about?</p> <p>A. a trip to the store</p> <p>B. the crowded sidewalks</p> <p>C. a toy shop</p> <p>D.</p>          | <p>37 On the sidewalks it was ____.</p> <p>A. cold</p> <p>B. hot</p> <p>C. wet</p> <p>D. windy</p>                              |
| <p>35 Benny and Al were in ____.</p> <p>A. a big city</p> <p>B. a small village</p> <p>C. the country</p> <p>D. the mountains</p>                | <p>38 Benny and Al went to the department that sold ____.</p> <p>A. books</p> <p>B. shoes</p> <p>C. tools</p> <p>D. toys</p>    |
| <p>36 The boys were in a hurry to get to the store because they were ____.</p> <p>A. excited</p> <p>B. hungry</p> <p>C. late</p> <p>D. tired</p> | <p>39 What do you think the boys will buy at the store?</p> <p>A. clothes</p> <p>B. glasses</p> <p>C. skates</p> <p>D. toys</p> |

Read the story. Answer the questions. Circle the letter of your answer.

Oh, how Peter wished he could whistle! Peter saw his friend Sam playing with a dog. Whenever Sam whistled, the dog ran straight to him. Peter wished he could do that trick with his own dog, Willie. Peter tried and tried to whistle, but he just couldn't.

Peter went into his house and put on his father's old hat, to make himself feel more grown-up. He looked into the mirror to practice whistling. Still no whistle!

|                                                                                                                                                                                    |                                                                                                                                                                             |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p>47</p> <p>What is the main idea of this story?</p> <p>A. Peter wishes he could whistle.</p> <p>B. Peter likes to play with his dog.</p> <p>C. Peter wears his father's hat.</p> | <p>50</p> <p>What trick did Peter want to do with his dog Willie?</p> <p>A. teach Willie to whistle</p> <p>B. put an old hat on Willie</p> <p>C. whistle to call Willie</p> |
| <p>48</p> <p>What happened when Sam whistled?</p> <p>A. Peter went over to see Sam.</p> <p>B. A dog went over to see Sam.</p> <p>C. Peter whistled, too.</p>                       | <p>51</p> <p>Why did Peter put on his father's hat?</p> <p>A. to see how he would look</p> <p>B. to make himself feel more grown-up</p> <p>C. to play a trick on Willie</p> |
| <p>49</p> <p>How many dogs are there in this story?</p> <p>A. one dog</p> <p>B. two dogs</p> <p>C. three dogs</p>                                                                  |                                                                                                                                                                             |

The next day Peter went outside to play. He sat on the front steps and tried to whistle. Then Peter saw his dog coming. Quick as a wink, Peter hid behind the stairs. He wanted to surprise Willie with a whistle. Peter puffed up his cheeks. He blew and blew and blew. Suddenly, out came a real whistle. Willie stopped and looked around to see who was making the noise.

"It's me," Peter shouted. He jumped out from behind the stairs. Willie raced straight up to him.

|                                                                                                                                                                                                                                                   |                                                                                                                                                                                                                                                                                                       |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <div data-bbox="203 642 240 672">52</div> <p>Where did Peter hide?</p> <ul style="list-style-type: none"> <li>A. under the bed</li> <li>B. on the front steps</li> <li>C. behind the stairs</li> </ul>                                            | <div data-bbox="863 642 900 672">54</div> <p>How did Peter feel when Willie came running?</p> <ul style="list-style-type: none"> <li>A. happy</li> <li>B. scared</li> <li>C. mad</li> </ul>                                                                                                           |
| <div data-bbox="203 1021 240 1050">53</div> <p>Why did Willie stop and look around?</p> <ul style="list-style-type: none"> <li>A. to see where Peter was</li> <li>B. to see the other dogs</li> <li>C. to see who was making the noise</li> </ul> | <div data-bbox="863 1021 900 1050">55</div> <p>What lesson can we learn from this story?</p> <ul style="list-style-type: none"> <li>A. Keep trying and you'll learn to do something.</li> <li>B. Every boy should have a dog</li> <li>C. If something is too hard, don't waste time on it.</li> </ul> |

Winter 1975  
Grade 5

NAME: \_\_\_\_\_ (LAST) \_\_\_\_\_ (FIRST)

Read the story on this page. Then answer the questions.

Most spies on TV today are men. Long ago, however, some of the most famous spies were women. One such female spy was Alice Dubois.

During World War I, France was attacked by Germany. Alice Dubois left her home in France and escaped to England. She told English officials many details that she knew about the German attack. The English officials thought she would be a good spy. They asked her to form a spy group. Soon Alice Dubois and her new spy group returned to France to get more information.

Alice Dubois worked out many methods of sending messages back to the English. She signaled with church bells. She hid messages in cakes and toys. She used invisible ink to write a message on a photograph. And once, she hid a map in the rim of her eye-glasses.

She was not entirely successful, though. German secret agents trailed her and finally arrested her. Her sentence was 27 years in prison. But she died in prison after four years, just a few weeks before the war ended.

1. This passage is mostly about \_\_\_\_\_.

- A. A female spy
- B. Spies on TV
- C. World War I
- D. Sending messages

2. Why did Alice Dubois leave France?

- A. France was attacked by its enemy.
- B. She wanted to visit England.
- C. She wanted to become a spy.
- D. She didn't want to be caught by the enemy.

3. In order to ask her to spy for them, the English officials must have thought Alice was \_\_\_\_\_.

- A. A trained spy
- B. Smart and clever
- C. A woman
- D. Easily frightened

4. Where did Alice Dubois send messages?
- A. to England
  - B. to Germany
  - C. to France
  - D. to the United States
5. What happened to Alice at the end of the story?
- A. Alice escaped from prison.
  - B. Alice was sent to prison just before the war ended.
  - C. Alice died in prison.
  - D. Alice watched spies on TV.
6. Which statement below is an opinion, rather than a fact?
- A. Alice Dubois was a female spy.
  - B. Alice Dubois was brave to be a spy.
  - C. Alice Dubois was sentenced to 27 years in prison.
  - D. Alice Dubois hid a message in a cake.

GO ON TO NEXT PAGE

Read the story carefully. Then answer the questions below.  
Circle the letter of the answer you choose.

Ida, who likes to draw, has four brothers. Archie and Ernie are her older brothers. They both like to read. Ernie likes to draw, too, although mostly he draws cowboy pictures. Sol is Ida's younger brother. He likes to do everything his older brother Ernie does. Ida likes to play with her brothers. Mostly she plays with her brothers who like to draw. On rainy days all the children stay inside and read, draw, or play games.

7. What does Sol like to do?

- A. read
- B. draw
- C. read and draw
- D. play games

8. How many of the children like to draw?

- A. 1
- B. 2
- C. 3
- D. 4

9. From reading the story, what do you think Archie probably does on a rainy day?

- A. read
- B. draw
- C. watch TV
- D. ride his bike

10. What kind of pictures do you think Sol draws most?

- A. racing car pictures
- B. pictures of trees
- C. cowboy pictures
- D. can't tell from the story

GO ON TO NEXT PAGE

Read the passage and then answer the questions that follow.

When the fourth grade class went to the zoo, they learned many things about gorillas. As the students watched, the gorillas began to play games just like those of human children. The gorillas wrestled with each other. They snake-danced up and down the slope as they held on to each other's tails. Then they played a game that resembled "follow the leader," which caused the youngest gorillas to make a giggling sound.

"These animals," the zookeeper said, "are called lowland gorillas." The zookeeper, who is an authority on gorillas, went on, "There are two types of gorillas. The lowland gorillas were discovered in 1856 by an American named Paul du Chaillu. These lowland gorillas are the ones most often seen in jungles and zoos. The other type of gorillas, the mountain gorilla, was first seen in 1902 by a German officer named Captain Oscar von Beringe.

"Our zoo--the St. Louis Zoo--is fortunate to have both types of gorillas. The zookeepers watch the habits of both types of gorillas to see how they are like or different from human habits. They have discovered that both types of gorillas can grow to be much larger than any human. For example, one of the lowland gorillas, which the zookeepers named Phil, grew to weigh about 600 pounds. To reach this enormous size, the gorillas are fed a diet that is similar to a human diet in terms of the type of food. However, the gorillas eat much larger amounts of food than most humans do. A typical day's menu for one gorilla consists of 20 bananas, 8 oranges, 6 apples, 1/2 grapefruit, 1 pineapple, 2 pears, 2 tomatoes, 2 carrots, 1 head of lettuce, 1/2 bunch of celery, 6 slices of bread, and 3 pints of a milk mixture.

"It is not surprising that after such a large meal the gorillas sleep for 12 to 15 hours before starting another day of 'human games' and 'gorilla eating'."

11. What is this passage mainly about?

- A. jungle animals
- B. gorillas
- C. a trip to the zoo
- D. the diet of a lowland gorilla

12. Which phrase describes the zookeeper?

- A. a German officer
- B. an American
- C. an expert on gorillas
- D. a resident of a jungle

GO ON TO NEXT PAGE

Here are some more questions about the last passage.  
Look back, if you need to, to check your answers.

13. What is the author's purpose in writing this story?
- A. to compare gorillas and humans
  - B. to describe the world's largest gorilla
  - C. to announce the discoveries of two types of gorillas
  - D. to name the playful activities of gorillas
14. What do the gorillas at the zoo do after they eat?
- A. wash themselves carefully
  - B. hibernate for several months
  - C. sleep for many hours
  - D. wrestle with each other
15. If you invited a gorilla to dinner, what would he enjoy most?
- A. meat and potatoes
  - B. soup
  - C. fruit salad
  - D. ice cream and cake
16. What is paragraph 2 telling about?
- A. gorilla games
  - B. an American explorer
  - C. two types of gorillas
  - D. the zookeeper's favorite animal
17. Which of the following is an example of an opinion rather than a fact?
- A. The gorillas probably sleep too much for their own good.
  - B. The mountain gorilla was first discovered in 1902.
  - C. The gorillas eat larger quantities of food than humans do.
  - D. The lowland gorillas are most often seen in jungles and zoos.
18. Which one is most qualified to write this selection?
- A. an actor in a TV show about gorillas
  - B. someone who has studied animals
  - C. the animal trainer in a circus
  - D. someone who has a chimpanzee for a pet

GO ON TO NEXT PAGE

Read the story and answer the questions.

One type of food with a really interesting name is the hot dog. At first this sausage was called a dachshund sausage, because it was long and thin, like the dachshund dog. The dachshund sausage got its new name when it was first served at a baseball game in the year 1900. On the day of the game the weather was cold. Food vendors knew that people would want something hot to eat. So they served dachshund sausages, which they kept hot in tubs of boiling water. The chilled baseball fans enjoyed the steaming sausages. So the vendors kept on serving dachshund sausages at baseball games. "Hot dachshunds," they would call, to tempt people to come and buy. But they had trouble spelling the word dachshund when they wanted to make signs. So pretty soon they changed the name, so it would be easier to spell. They tried different names, but the one that stuck was--hot dog. That's how the hot dog got its name.

19. This story is mainly about:

- A. what an interesting name the hot dog has.
- B. how the hot dog got its name.
- C. when hot dogs were first eaten.
- D. why hot dogs are sold at baseball games.

20. What was the sausage called at first?

- A. hot dog
- B. hot sausage
- C. thin dog
- D. dachshund sausage

21. According to the story, why did the baseball fans enjoy the sausages?

- A. because they didn't cost much
- B. because they were hot
- C. because they tasted good
- D. because the sausages came from a foreign country

22. Why was the name changed?

- A. so it would be easier to say
- B. so it would be easier to spell
- C. so it would look better on signs
- D. so people could remember it

GO ON TO NEXT PAGE

Read the poem and then answer the questions.

Racing his engines that fateful day  
Casey Jones and his crew thought they had a clear way.  
But just up ahead, past Hoboken,  
Sat two freight trains, their engines broken.

Casey saw them in the nick of time  
(His engines were racing at ninety-nine).  
He pulled on the whistle and yanked on the brakes  
And hoped that he had the strength that it takes

To stop a freight train (roaring so fast  
To meet a deadline two miles past  
The broken trains that sat up ahead)-  
Casey Jones was filled with dread.

With a screech and a roar and a large puff of smoke  
The brakes hit the wheels and suddenly broke!  
And now when people discuss Casey Jones,  
They lower their voices in reverent tones.

-- Nerak Fozreh

- |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p>23. What is this poem about?</p> <ul style="list-style-type: none"> <li>A. a trip to Hoboken</li> <li>B. racing cars</li> <li>C. freight trains</li> <li>D. a train accident</li> </ul> <p>24. Which one happened right after Casey yanked on the brakes?</p> <ul style="list-style-type: none"> <li>A. The train stopped.</li> <li>B. The train left the station.</li> <li>C. Casey spotted the disabled trains.</li> <li>D. The train screeched, howled, and smoked.</li> </ul> | <p>25. How did Casey feel when he pulled on the brakes?</p> <ul style="list-style-type: none"> <li>A. confident, sure of himself</li> <li>B. weak</li> <li>C. afraid</li> <li>D. happy and excited</li> </ul> <p>26. Why do people lower their voices when they discuss Casey Jones now?</p> <ul style="list-style-type: none"> <li>A. They are afraid of him.</li> <li>B. The brakes broke.</li> <li>C. Casey Jones won the race.</li> <li>D. Casey Jones died.</li> </ul> |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

GO ON TO NEXT PAGE

Read the passage and answer the questions on the next 2 pages.

The Etruscans were an interesting group of people. Historians believe they lived in Italy even before the fifth century B.C., 2,500 years ago. Not only are these people well known for building cities in honor of their dead ancestors, but they are also known for other interesting customs.

One unusual custom of the Etruscans was their way of hunting for food. They did use the usual nets and dogs to catch deer and bears, but they also used music. First, they would set out nets and traps for the animals. Next, a person would play a pleasing melody on a flute. When the animals heard this melody, they seemed to forget about their safety. They would go toward the music and be caught in the traps.

Another interesting custom of the Etruscans was that of creating beautiful objects to bury with their dead. Such things as bronze figures, ceramic statues, beautifully decorated gold jewelry, vases, mirrors, and paintings were often found in or on the tombs of the ancient Etruscans.

Archaeologists, who study about ancient peoples, think there are many other interesting customs of the Etruscans. They continue in their search to find out more about these interesting people.

27. Which sentence in the story best states the main idea?

- A. the first sentence
- B. the second sentence
- C. the third sentence
- D. the fourth sentence
- E. the fifth sentence

28. How long ago did the Etruscans live?

- A. 5 years ago
- B. 10 centuries ago
- C. 2,000 years ago
- D. 2,000 centuries ago
- E. 2,500 years ago

29. Which person is most qualified to write this article?

- A. an artist
- B. a historian
- C. an archeologist
- D. both A. and B.
- E. both B. and C.

30. What is paragraph 2 about?

- A. archeologists
- B. Etruscan art objects
- C. hunting for food
- D. Etruscan music
- E. ancient history

31. What did the Etruscans use to catch animals?

- A. bows and arrows
- B. guns
- C. music
- D. mirrors
- E. their bare hands

32. Which school subject would an Etruscan probably like best?

- A. English
- B. history
- C. mathematics
- D. gym
- E. art

GO ON TO NEXT PAGE

33. According to the passage, why did the Etruscans build great cities?

- A. to provide housing for all their people
- B. to honor dead ancestors
- C. to honor the ruler
- D. because there were so many traders passing through
- E. because they had beautiful marble for buildings

34. Where would the flute player stand?

- A. right next to the trap
- B. a short distance behind the trap
- C. with the hunting dogs
- D. with the other musicians
- E. where he could lead the animals away from the trap

35. Which of the following is an example of a fact rather than an opinion?

- A. The Etruscans built cities long ago.
- B. The jewelry made by the Etruscans was the most beautiful ever made.
- C. Historians do not know as much as archaeologists do.
- D. The Etruscans had a very clever way of hunting for food.
- E. The customs of the Etruscans are more interesting than those of early Americans

GO ON TO NEXT PAGE

Read the story, then answer the questions on the next page.



When I slam out of our house mad, I go along over to my Aunt Kate's. She's not really my aunt. The kids around here call her Crazy Kate the Cat Woman because she walks along the street in funny old clothes and sneakers talking to herself, and she sometimes has half a dozen or more stray cats living with her. I guess she does sound a little looney, but it's just because she does things her own way, and she doesn't give a hoot what people think. She's sane, all right. In fact, she makes a lot better sense than my pop.

It was three or four years ago, when I was a little kid, and I came tearing down our stairs crying mad after some fight with Pop, that I first met Kate. I plunged out of our door and into the street without looking. At the same moment I heard brakes scream and felt someone yank me back by the scruff of my neck. I got dropped in a heap on the sidewalk.

I looked up, and there was a shiny black car with M.D. plates and Kate waving her umbrella at the driver and shouting: "Listen, Dr. Big Shot, whose life are you saving? Can't you even watch out for a sniveling kid crossing the street?"

The doctor looked pretty sheepish, and so did I. A few people on the sidewalk stopped to watch and snicker at us. Our janitor Butch was there, shaking his finger at me. Kate nodded to him and told him she was taking me home to mop me up.

GO ON TO NEXT PAGE

Read each question and circle the letter of the answer you choose.  
Look back at the story if you want to.

36. What is the author mainly doing in this story?

A. telling about a funny incident  
B. building up suspense  
C. describing where the boy lives  
D. introducing Kate  
E. setting an unhappy mood

37. The boy usually goes to Kate's when he is \_\_\_\_\_.

A. angry  
B. sad  
C. sick  
D. bored  
E. happy

38. The description of Kate suggests that she \_\_\_\_\_.

A. has many friends  
B. has a mind of her own  
C. has a lot of money  
D. speaks softly and kindly  
E. takes pride in how she looks

39. Which of the following is the boy's belief, rather than a fact?

A. Kate is not really my aunt.  
B. The kids call her Crazy Kate.  
C. She makes better sense than my father.  
D. I met her after a fight with Pop.  
E. She picked me up by the scruff of my neck.

40. Where does the main action of the story take place?

A. on a city sidewalk  
B. in front of the boy's house  
C. on a staircase  
D. on a country road  
E. in a small town

41. She doesn't give a hoot.  
What does this sentence mean?

A. She won't give up any cause.  
B. She never talks about others.  
C. She's not concerned about it.  
D. She doesn't scream or make noises.  
E. She never gives a reasonable answer.

42. When Kate "mopped up" the boy, she

A. cleaned him  
B. picked him up  
C. brushed his teeth  
D. knocked him down  
E. wiped him with a mop

APPENDIX D

Group Testing Materials for Grade 5 Mathematics

(Study 4)

**21**

Spring 1976

Individualized  
Group Pretest  
Grade 5 Math

NAME: \_\_\_\_\_

Last

First

1

Mary, Jenny, and Joan had a bake sale. Mary brought 48 cookies. Jenny and Joan brought 62 each. How many did they bring in all?

- A. 110 cookies      C. 200 cookies  
B. 152 cookies      D. 172 cookies

3

A troop of 19 Girl Scouts joined another troop of 13 for a hike. They walked in rows of four. How many rows did they make?

- A. 7                      C. 9  
B. 8                      D. 6

2

The Danube River is 1770 miles long and the MacKenzie River is 2635 miles long. How much longer is the MacKenzie River than the Danube River?

- A. 865 miles      C. 4,405 miles  
B. 1,965 miles      D. 965 miles

4

Anne has 12 dimes. George has 5 dimes less than Anne. Which equation can be used to tell how many dimes they have altogether?

- A.  $12 - 5 = \square$   
B.  $12 + 5 = \square$   
C.  $12 + (12 - 5) = \square$   
D.  $(12 + 5) + 12 = \square$

5

Janet had \$4.00 to repair her bicycle. She paid \$2.25 for a tire, \$.75 for a seat, and \$.50 for paint. How much did she have left?

- A. \$.25                      C. \$1.00  
B. \$.50                      D. \$1.50

6

Last Friday Jeff bowled games scoring 121, 99, 154, and 110. What was his average score?

- A. 120                      C. 242  
B. 121                      D. 142

Questions 7 and 8 are to be answered using this word problem:

During Janet's last vacation, she and her family drove to Oregon. They averaged 50 miles an hour and got 25 miles per gallon of gas. The trip took them 9 hours.

7

How many miles did they drive?

- A. 425 miles                      C. 405 miles  
B. 225 miles                      D. 450 miles

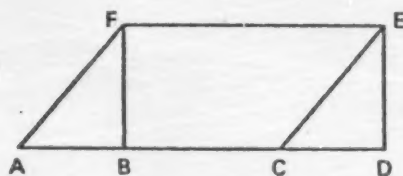
8

How many gallons of gas did they use during the trip?

- A. 18 gallons                      C. 20 gallons  
B. 14 gallons                      D. 16 gallons

9

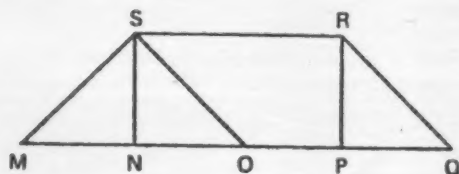
In the figure below, what is the name of the geometric figure AFEC?



- A. Rectangle
- B. Square
- C. Triangle
- D. Parallelogram

10

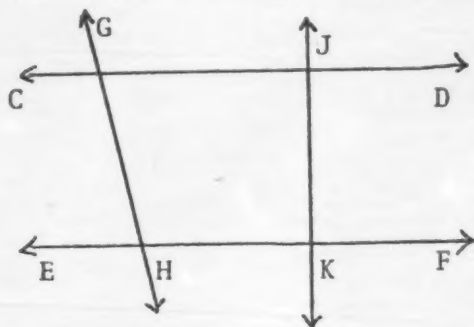
Which of the following names a rectangle?



- A. SRPO
- B. MSO
- C. SRPN
- D. SRQO

11

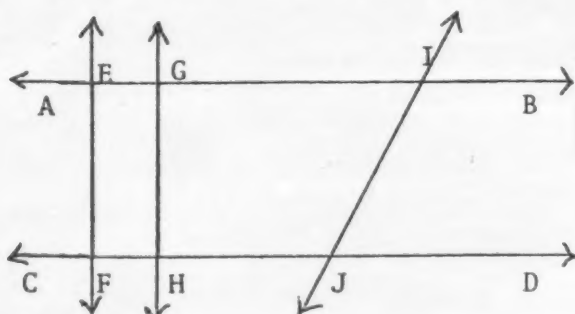
Which pair of lines is perpendicular?



- A.  $\overleftrightarrow{EF}$  and  $\overleftrightarrow{CD}$
- B.  $\overleftrightarrow{JK}$  and  $\overleftrightarrow{GH}$
- C.  $\overleftrightarrow{GH}$  and  $\overleftrightarrow{EF}$
- D.  $\overleftrightarrow{JK}$  and  $\overleftrightarrow{CD}$

12

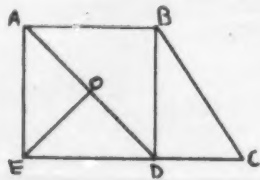
Which pair of lines is parallel?



- A.  $\overleftrightarrow{EF}$  and  $\overleftrightarrow{GH}$
- B.  $\overleftrightarrow{EF}$  and  $\overleftrightarrow{CD}$
- C.  $\overleftrightarrow{GH}$  and  $\overleftrightarrow{IJ}$
- D.  $\overleftrightarrow{AB}$  and  $\overleftrightarrow{IJ}$

3

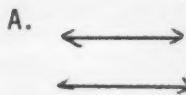
Which of the following names a square?



- A. EDA                      C. ABCD  
B. ABDE                    D. ABCE

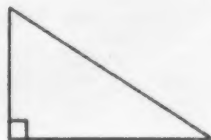
17

Which figure shows perpendicular lines?



14

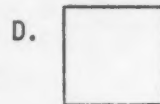
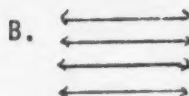
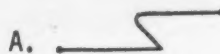
This figure is a \_\_\_\_?



- A. Obtuse triangle          C. Parallelogram  
B. Rhombus                D. Right triangle

18

Which figure contains 4 line segments?



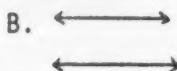
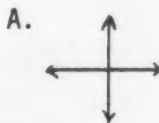
15

Which one of these is an angle?



19

Which figure shows parallel lines?



16

Which one of these is a line segment?



20

What is the name of this figure?



A. Line

C. Chord

B. Ray

D. Line segment

21

 $\begin{array}{r} 312 \\ \times 3 \\ \hline \end{array}$ 

A. 1036

B. 938

C. 946

D. 936

25

 $\begin{array}{r} 243 \\ \times 4 \\ \hline \end{array}$ 

A. 862

B. 872

C. 972

D. 982

22

 $\begin{array}{r} 9423 \\ \times 2 \\ \hline \end{array}$ 

A. 18,946

B. 19,856

C. 18,846

D. 1,846

26

 $\begin{array}{r} 12 \\ \times 44 \\ \hline \end{array}$ 

A. 528

B. 428

C. 96

D. 86

23

 $\begin{array}{r} 15 \\ \times 5 \\ \hline \end{array}$ 

A. 75

B. 55

C. 65

D. 85

27

 $\begin{array}{r} 87 \\ \times 14 \\ \hline \end{array}$ 

A. 1218

B. 1198

C. 445

D. 1118

24

 $\begin{array}{r} 37 \\ \times 6 \\ \hline \end{array}$ 

A. 182

B. 222

C. 202

D. 228

28

 $\begin{array}{r} 121 \\ \times 32 \\ \hline \end{array}$ 

A. 3705

B. 1872

C. 605

D. 3872

29

 $\begin{array}{r} 193 \\ \times 79 \end{array}$ 

- A. 3,088
- B. 14,247
- C. 15,247
- D. 15,047

33

 $\begin{array}{r} 49 \\ \times 52 \end{array}$ 

- A. 2148
- B. 2548
- C. 343
- D. 2448

30

 $\begin{array}{r} 32517 \\ \times 4 \end{array}$ 

- A. 128,068
- B. 130,068
- C. 120,068
- D. 140,068

34

 $\begin{array}{r} 278 \\ \times 30 \end{array}$ 

- A. 834
- B. 8380
- C. 6140
- D. 8340

31

 $\begin{array}{r} 43 \\ \times 32 \end{array}$ 

- A. 215
- B. 1276
- C. 1376
- D. 205

35

 $\begin{array}{r} 402 \\ \times 16 \end{array}$ 

- A. 2814
- B. 4272
- C. 6432
- D. 6834

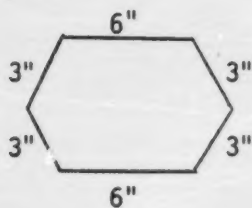
32

 $\begin{array}{r} 65 \\ \times 24 \end{array}$ 

- A. 1560
- B. 1540
- C. 1460
- D. 390

36

What is the perimeter of this polygon?



A. 45 inches

B. 36 inches

C. 18 inches

D. 24 inches

37

What is the perimeter of a square whose sides are each 8 inches?

A. 64 inches

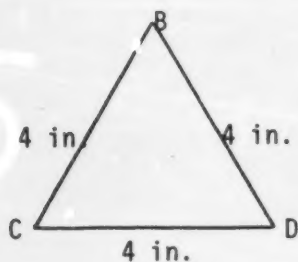
B. 32 inches

C. 24 inches

D. 22 inches

38

What is the perimeter of triangle BCD?



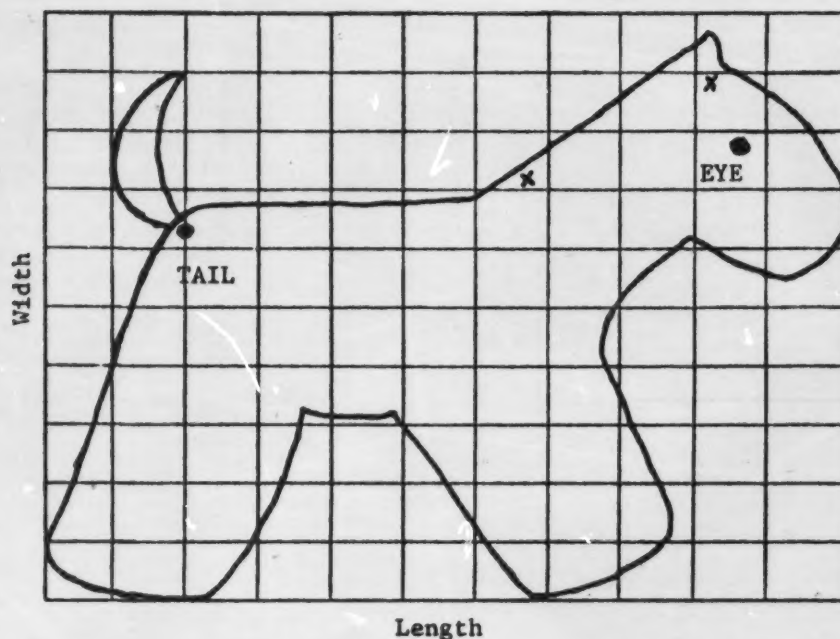
A. 12 inches

B. 16 inches

C. 20 inches

D. 18 inches

## POLKA DOT PONY



SCALE: Each Square is 2 inches long and 2 inches wide.

Here is a pattern for making an animal. Following this pattern, what size piece of paper would you use to make one animal?

39. The paper's width would be

- A. 20 inches    B. 22 inches    C. 10 inches    D. 11 inches

40. The paper's length would be

- A. 11 inches    B. 22 inches    C. 10 inches    D. 20 inches

41. The paper's perimeter would be

- A. 42 inches    B. 40 inches    C. 84 inches    D. 480 inches

## WORD PROBLEMS

| # | Rdg. | Trbl Sel<br>Opertn | Used App.<br>Info. | Selt. Corr.<br>Opertn. | Comp<br>Errors | Incl.all<br>info c̄ prmt | Vocabulary Prob<br>- Notes - |
|---|------|--------------------|--------------------|------------------------|----------------|--------------------------|------------------------------|
| 1 |      |                    |                    |                        |                |                          |                              |
| 2 |      |                    |                    |                        |                |                          |                              |
| 3 |      |                    |                    |                        |                |                          |                              |
| 4 |      |                    |                    |                        |                |                          |                              |
| 5 |      |                    |                    |                        |                |                          |                              |
| 6 |      |                    |                    |                        |                |                          |                              |
| 7 |      |                    |                    |                        |                |                          |                              |
| 8 |      |                    |                    |                        |                |                          |                              |

Additional WP Notes:

|       | OPEN ENDED                     | TEST ITEMS | Prob. c̄ Ltr. Refer. | Notes |
|-------|--------------------------------|------------|----------------------|-------|
| 9     |                                |            |                      |       |
| 10    |                                |            |                      |       |
| 11/17 |                                |            |                      |       |
| 12/19 |                                |            |                      |       |
| 13    |                                |            |                      |       |
| 14    |                                |            |                      |       |
| 15    |                                |            |                      |       |
| 16/18 |                                |            |                      |       |
| 20    |                                |            |                      |       |
| 37    | Definition of perimeter: F G E |            |                      |       |
| 39    | USED SCALE conversion Y N      |            |                      |       |
| 40    | Used scale conversion Y N      |            |                      |       |
| 41    | 41. Counted Boxes Y N          |            |                      |       |
|       | 41. Counted Edges Y N          |            |                      |       |
|       | 41. Computed Ans. Y N          |            |                      |       |

Tendency to ignore scale information Y N